
**Evaluating the impact of a tablet-based
intervention on the mathematics attainment,
receptive language and approaches to learning
of preschool children**

Jodie Walton

Thesis submitted to the University of Nottingham
for the degree of Doctorate in Applied Educational
Psychology

September 2018

Abstract

Hand-held technology is increasingly being used in educational settings as a medium of instruction for young children (Hubber et al., 2016). Although the evidence base is developing, little is currently known about the effectiveness of mathematics interventions delivered through tablet technology, particularly for preschool children in the UK. The present research evaluates the impact of the *onebillion* tablet-based intervention on the mathematics attainment, receptive language and positive ‘approaches to learning’ of 3-4 year old children.

An embedded mixed methods design was used in this study. The primary aims of the research were addressed through a quasi-experimental, ability-matched design. Across two nurseries, forty-seven children were allocated to either an experimental group, who accessed the intervention for fifteen minutes per day over 9 weeks ($n = 23$), or a control group ($n = 24$). Additional nested data was collected, including qualitative semi-structured facilitator interviews and observations, to further illuminate factors affecting outcomes.

At post-test, the experimental group had significantly higher mathematics attainment than the control group (controlling for pre-test ability), assessed on a researcher-developed measure of curriculum knowledge. At 5 month follow-up, the experimental group still appeared to outperform children in the control group, but differences between groups were no longer statistically significant. There was no significant intervention effect on a standardised measure of mathematics, or other aspects of development, including children’s receptive language or ‘approaches to learning’.

Based upon analysis of embedded data, a model is proposed of the potential mechanisms underpinning the efficacy of the intervention, accounting for individual differences and implementation factors on outcomes.

Findings from this study are discussed in relation to relevant literature and theory. Methodological limitations of the study are also acknowledged, as well as the implications of these findings for the use of educational technology in the early years, the practice of educational psychologists and further research.

Acknowledgements

There are a number of people I would like to thank for their support in completing this research and my doctoral training:

- To the staff, parents and children who participated in this research, particularly the teaching staff, for their commitment and hard work which made this project possible.
- To the tutor team at Nottingham University, especially Anthea Gulliford for her fantastic encouragement, guidance and supervision in conducting this project and more widely across the doctorate programme.
- To Nicola Pitchford and the rest of the Nottingham research team for sharing their knowledge in this field and supporting the development of this work.
- To the not-for-profit organisation, *onebillion*, for providing the app software free of charge to schools participating in this research.
- To the wonderful team at my placement Educational Psychology Service, for their kindness and support which have made my training so enjoyable - special thanks to Vanessa for her time and investment in my professional development.
- To my fellow trainees at the University of Nottingham for your infectious enthusiasm, thoughtfulness and friendship which have supported me on this journey. I cannot name you all individually but I have learnt so much from each of you.
- To my friends and family for their love and encouragement, especially to my parents, Phil and Karen, for their unwavering support, and to my mother-in-law, Anne, for her proof-reading skills!
- Finally, to my husband Ben, for his endless patience, sense of humour and belief in me – thank you.

Table of Contents

1 Introduction	1
1.1 Focus of the Research	1
1.2 Background of the Research	1
1.3 Overview of Chapters	2
2 Literature Review	4
2.1 ‘Closing the gap’ in mathematics achievement in the UK	4
2.2 The importance of early years education	6
2.2.1 Typical development of early mathematical skills	6
2.2.2 Early mathematics as a foundation for later success	9
2.2.3 Individual differences in preschool mathematics attainment	9
2.3 Early Intervention in Mathematics	10
2.3.1 What are the features of effective early intervention in mathematics?	11
2.3.2 How effective are current early years mathematics interventions?	13
2.4 The Role of Technology in Early Mathematics Intervention	16
2.4.1 How can the use of tablet technology benefit young children’s learning in mathematics?	17
2.4.2 What impact does educational technology have on mathematical attainment?	17
2.4.3 Criticisms regarding the use of educational technology with young children	19
2.4.4 What aspects of implementation might affect learning outcomes from educational technology?	20
2.5 The <i>Onebillion</i> Tablet-Based Mathematics Intervention	21

2.5.1	What features of the <i>onebillion</i> software may support mathematical development?	21
2.5.2	How effective is the <i>onebillion</i> tablet intervention?	23
2.6	Systematic Review of the Evidence-Base	23
2.6.1	Systematic review methodology	24
2.6.2	Eligibility criteria	24
2.6.3	Database search	25
2.6.4	Screening	26
2.6.5	Quality and relevance of research	28
2.6.6	Critical analysis of the evidence-base	28
2.6.7	Synthesis of key findings	36
2.6.8	Overall summary of the evidence-base	40
2.7	The Potential Cross-Domain Impact of Tablet Interventions	42
2.7.1	Language development	42
2.7.2	Approaches to learning	44
2.8	The Aims and Original Contribution of the Proposed Research	47
2.8.1	Raising mathematical attainment in the preschool years	47
2.8.2	The cross-domain impact of the <i>onebillion</i> intervention	47
2.8.3	Factors affecting intervention outcomes	47
2.9	Research Questions	48
3	Methodology	49
3.1	Paradigms within Applied Educational Research	49
3.1.1	Alternative paradigms	49
3.1.2	The philosophical assumptions underpinning the current study ...	50
3.2	Stakeholders	51
3.3	Mixed Methods Research Design	52
3.3.1	Mixed methods and applied research	52

3.3.2	Research design of the current study	54
3.4	Quantitative Evaluation of the Efficacy of the Intervention	57
3.4.1	Research Questions 1-4 and associated hypotheses	57
3.4.2	Evaluation design	59
3.4.3	Sampling.....	60
3.4.4	Conditions.....	65
3.4.5	Measures	66
3.4.6	Procedure	74
3.4.7	Validity and reliability	76
3.5	Embedded Aspects of the Research Design.....	82
3.5.1	Research Questions 5 and 6	82
3.5.2	Semi-structured facilitator interviews	82
3.5.3	Narrative observations	85
3.5.4	Structured observations of attention on-task.....	86
3.5.5	Associations between learner characteristics and intervention gains	87
3.6	Ethical Considerations	88
3.7	Summary.....	93
4	Results.....	94
4.1	Quasi-Experimental Evaluation Data Analysis	94
4.1.1	Data analysis procedures	94
4.1.2	Data analysis for Research Question 1.....	98
4.1.3	Data analysis for Research Question 2.....	105
4.1.4	Data analysis for Research Question 3.....	109
4.1.5	Data analysis for Research Question 4.....	111
4.1.6	Summary of quasi-experimental findings	113
4.2	Embedded Data Analysis	113

4.2.1	Qualitative data analysis procedures	114
4.2.2	Data analysis for Research Questions 5 and 6	118
4.2.3	Research Question 5: Key themes	121
4.2.4	Research Question 6: Key themes	125
4.2.5	Quantitative data analysis for Research Question 6	140
4.2.6	Chapter summary	144
5	Discussion	145
5.1	Research Findings and Theoretical Relevance	145
5.1.1	Outcomes of the intervention for mathematical attainment	145
5.1.2	The longitudinal impact of the intervention	149
5.1.3	Outcomes of the intervention for receptive language skills	151
5.1.4	Outcomes of the intervention for approaches to learning	152
5.1.5	Factors affecting the outcomes of the intervention	153
5.2	Methodological Limitations	160
5.2.1	Limitations of the quasi-experimental design	160
5.2.2	Limitations of embedded aspects of the design	163
5.2.3	Ethical considerations	165
5.3	Research Implications	166
5.3.1	Implications for the use of tablet technology within early years education	166
5.3.2	Implications for the future development of tablet-based early years mathematics interventions	167
5.3.3	Implications for the professional work of educational psychologists	168
5.3.4	Directions for future research	169
5.4	Additional Reflections	171
6	Conclusions	173
7	References	175

8 Appendices	203
8.1 Systematic Literature Review: Reasons for excluding articles following title and abstract screening	203
8.2 Systematic Literature Review: Excluded full-text articles	204
8.3 Weight of Evidence Criteria Judgements for Systematic Literature Review	205
8.4 Systematic Literature Review: Comparison of Included Study Features	209
8.5 Expression of Interest Letter	218
8.6 Head Teacher Information Sheet and Consent Form	220
8.7 Parent Information Sheet and Consent Form	225
8.8 Child 'Willingness to Participate' Form	230
8.9 Curriculum Knowledge Assessment Example of Questions and Answer Booklet (from Outhwaite et al., 2017)	231
8.10 Facilitator Training	233
8.11 Intervention Fidelity Check	235
8.12 Facilitator Information Sheet and Consent Form	237
8.13 Semi-Structured Facilitator Interview Script	240
8.14 Debriefing Procedures	243
8.15 Ethical Approval	246
8.16 Preliminary Checks for Analysis of Covariance	248
8.17 Thematic Analysis Step 1: Transcription Notation System	254
8.18 Thematic Analysis Step 2: Example of Initial Data Coding from Anne's Interview Transcript	255
8.19 Thematic Analysis Step 2: List of Final Codes from Interviews	257
8.20 Thematic Analysis Step 2: List of Final Codes from Narrative Observations	259
8.21 Thematic Analysis Step 3: Initial Thematic Maps of Interview Data	261

8.21.1	Research Question 5: What are facilitators' perceptions of the outcomes of the <i>onebillion</i> intervention for preschool children aged 3-4 years? ..	261
8.21.2	Research Question 6: What factors may affect the outcomes of the <i>onebillion</i> intervention for preschool children aged 3-4 years? ..	262
8.22	Thematic Analysis Step 3: Initial Thematic Map of Narrative Observational Data ..	264
8.22.1	Research Question 6: What factors may affect outcomes from the <i>onebillion</i> intervention for preschool children aged 3-4 years? ..	264
8.23	Thematic Analysis Step 4: An Example of Inter-Rater Reliability Checks for TA ..	266
8.24	Comparison between Intervention Topics and BAS-III Early Number Concepts Assessment ..	268

List of Figures

Figure 2.1: The Instructional Hierarchy (adapted from Haring & Eaton, 1978; Martens & Witt, 2004)	15
Figure 2.2: The process of screening and selecting records for the systematic review (adapted from Moher, Liberati, Tetzlaff, Atman & the PRISMA Group, 2009)	27
Figure 3.1: The Qualitative-Mixed Methods-Quantitative (QUAL-MM-QUAN) continuum (adapted from Teddlie and Tashakkori, 2009, p.28)	53
Figure 3.2: An illustration of the embedded mixed methods design	56
Figure 3.3: Pupil sampling flowchart	63
Figure 4.1: A line graph to show change in experimental and control conditions over time in Early Number Concept ability scores.....	100
Figure 4.2: A line graph to show change in experimental and control conditions at T1 and T2 in Curriculum Knowledge raw scores.....	103
Figure 4.3: A line graph to show change in experimental and control conditions at T1, T2 and T3 in Curriculum Knowledge raw score	107
Figure 4.4: Thematic map for Research Question 5: What are facilitators' perceptions of the outcomes of the intervention? The number of data extracts relevant to each theme/subtheme is given in italics.	122
Figure 4.5: Thematic map for Theme 1: Language demands of the intervention subthemes	126
Figure 4.6: Thematic map for Theme 2: Children's attention on-task subthemes	128
Figure 4.7: Thematic map for Theme 3: Pedagogy and instructional level of the intervention subthemes	131
Figure 4.8: Thematic map for Theme 4: Children's attitude to learning subthemes	134
Figure 4.9: Thematic map for Theme 5: Implementation in the nursery setting subthemes	138
Figure 4.10: A bar chart displaying children's attention on-task during intervention sessions, by group and week of the intervention	141

Figure 5.1: A model of the mechanisms underpinning the efficacy of the onebillion tablet intervention	155
--	-----

List of Tables

Table 2.1: Developmental milestones in counting, number and shape (summarised from Clements & Sarama, 2009)	8
Table 2.2: Effect sizes from meta-analyses evaluating the impact of educational technology on children's mathematics attainment.....	18
Table 2.3: Eligibility criteria for studies to be included in the systematic review	24
Table 2.4: Search terms included in the systematic review	25
Table 2.5: Database search results (search conducted 06/04/18)	25
Table 2.6: Age distribution of participants across studies included in the systematic review.....	30
Table 2.7: Key features of interventions included in the systematic review	33
Table 3.1: Research questions and hypotheses for the quasi-experimental design	58
Table 3.2: Demographic details of participating schools	62
Table 3.3: Demographic details of participating sample.....	64
Table 3.4: Typical mathematics instruction across settings	66
Table 3.5: Standardised mathematics measures for 3-4 year old children	68
Table 3.6: Standardised receptive language measures for 3-4 year old children	71
Table 3.7: Approach to learning teacher rating scales for children aged 3-4 years	73
Table 3.8: Delivery of the onebillion intervention for the experimental condition	75
Table 3.9: Internal validity of the research design.....	77
Table 3.10: Intervention fidelity checks	80
Table 3.11: Threats to reliability	81
Table 3.12: Facilitator details, including confidence, attitude and experience in using technology	84
Table 3.13: Ethical considerations applicable to the present research	89
Table 4.1: Descriptive statistics for Early Number Concepts ability scores across experimental and control conditions at T1 and T2.....	99

Table 4.2: Effect size analysis of the difference between conditions in T2 Early Number Concepts ability scores	101
Table 4.3: Descriptive statistics for Curriculum Knowledge scores for experimental and control conditions at T1 and T2.....	102
Table 4.4: Effect size analysis of the difference between conditions in T2 Curriculum Knowledge scores	104
Table 4.5: Descriptive statistics for the mathematics Curriculum Knowledge scores for experimental and control conditions at T1, T2 and T3 for the reduced sample.....	106
Table 4.6: Effect size analysis of the difference between conditions in T3 mathematics Curriculum Knowledge.....	108
Table 4.7: Descriptive statistics for receptive language for experimental and control conditions at T1 and T2.....	110
Table 4.8: Descriptive statistics for the approaches to learning for experimental and control conditions at T1 and T2.....	112
Table 4.9: Quality and trustworthiness of qualitative aspects of the design	116
Table 4.10: Narrative observations conducted during intervention sessions ..	118
Table 4.11: Correlational analyses between learning gains on the mathematics Curriculum Knowledge test and learner characteristics	143

List of Abbreviations

AAP	American Academy of Pediatrics
ANCOVA	Analysis of Covariance
AtL	Approaches to Learning
BAS	British Ability Scales
BPS	British Psychological Society
CELF-P2	Clinical Evaluation of Language Fundamentals – Preschool 2
CK	Curriculum Knowledge
CPA	Cognitive-Pictorial-Abstract
DECA-P2	Devereux Early Childhood Assessment for Preschoolers 2 nd edition
DfE	Department for Education
DV	Dependent Variable
EAL	English as an Additional Language
EEF	Education Endowment Foundation
ENC	Early Number Concepts
EP	Educational Psychologist
EPS	Educational Psychology Service
ES	Effect Size
EYFS	Early Years Foundation Stage
HCPC	Health Care and Professions Council
HLTA	Higher-Level Teaching Assistant
IDACI	Income Deprivation Affecting Children Index
IV	Independent Variable
IQR	Interquartile Range
MM	Mixed Methods

NCETM	National Centre for Excellence in the Teaching of Mathematics
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
RCT	Randomised Control Trial
QUAL	Qualitative
QUAN	Quantitative
RtI	Response to Intervention
SD	Standard Deviation
SEND	Special Educational Needs and Disabilities
SES	Socio-Economic Status
SIT	Social Identity Theory
SLR	Systematic Literature Review
TA	Thematic Analysis
TAU	Treatment as Usual
TEP	Trainee Educational Psychologist
T1	Time 1
T2	Time 2
T3	Time 3
UoN	University of Nottingham
US	United States
WoE	Weight of Evidence
ZPD	Zone of Proximal Development

1 Introduction

1.1 Focus of the Research

The purpose of this study was to evaluate the efficacy of a mathematics tablet-based intervention for preschool children, aged 3-4 years. An intervention developed by a not-for-profit organisation, *onebillion*, was implemented in nursery classrooms as a supplementary form of instruction alongside children's typical mathematics curriculum. This study primarily aimed to determine the effect of the intervention on children's attainment in mathematics, whilst also considering its possible impact on other areas of development, including children's language skills and approaches to learning in the classroom.

1.2 Background of the Research

The author is currently a Trainee Educational Psychologist (TEP), studying at the University of Nottingham (UoN), and on a training placement with a Local Authority Educational Psychology Service (EPS). The author's interest in the development of mathematics stemmed from her undergraduate studies in psychology, when she undertook research exploring the relationship between attention and early number skills with preschool children. Later, whilst working as a primary school teacher with 7-9 year old children, the author noted the large gap in attainment between children of this age range within the school, and felt that early, effective intervention could be beneficial in promoting not only higher attainment, but also more positive attitudes to mathematics.

During her time on placement as a TEP, the author was involved in a number of casework activities where children were experiencing difficulties in learning mathematics, including some children in the early years of school education. It was also evident that many teachers were looking for more effective ways of teaching mathematics and raising achievement for all. Whilst studying at the UoN, the author was introduced to the *onebillion* tablet intervention by a team of researchers who had already evaluated its impact on mathematics with older children and in other countries (Outhwaite, Gulliford & Pitchford, 2017; Outhwaite, Faulder, Gulliford & Pitchford, in press; Pitchford, 2015). There were

anecdotal reports that younger, 3-4 year old children were able to access and benefit from using the apps, but no formal evaluation had taken place. The author felt that further research with the *onebillion* tablet intervention might be fruitful in supporting early intervention, addressing gaps in attainment prior to school entry.

The research team at the UoN had also begun exploration of gains beyond mathematics attainment (e.g. attention, motor skills) (Pitchford & Outhwaite, in prep.). It was reported that teachers involved in past studies had commented, informally, that the app appeared to support children's understanding of mathematical language and instructions, as well as some children's persistence in problem-solving and confidence. This prompted the author to further investigate the impact of the intervention on domains outside of the area of mathematics within the preschool age group, including children's receptive language and 'approaches to learning' in the classroom.

The focus of this research is highly relevant to the work of educational psychologists (EPs) who have an important role in promoting evidence-based practice in educational settings, including nurseries, and advocating for early intervention. The mechanisms underpinning the efficacy of the *onebillion* intervention for preschool children are also explored in this study, with implications for EPs in guiding school implementation of educational technology in the nursery classroom.

1.3 Overview of Chapters

In the next chapter, the current political and educational context of the research is considered, highlighting the need for greater early intervention in mathematics in the UK. The potential for educational technology to raise attainment is then discussed, with reference to relevant psychological theory and evidence. A systematic literature review of previous research evaluating tablet-based mathematics interventions for young children is also presented, providing a rationale for the key focuses of the present research.

Chapter 3 presents the methodology of this study, sharing the researcher's philosophical positioning and the reasons for using mixed methods to evaluate

this intervention. The specific quantitative and qualitative methods are clearly outlined to support interpretation of the findings and future replication of the study.

In Chapter 4, procedures for data analysis of both the main quasi-experimental study and embedded aspects of the research design are explained. Results of both elements of the research are then presented.

The key findings of this research are discussed in Chapter 5, in relation to previous research and theory. Limitations are also addressed, as well as the implications of this study for the use and development of educational technology in early years education and for the professional work of EPs. Directions for further research are suggested.

Finally, Chapter 6 presents the key conclusions that can be drawn from this research and the original contribution these findings add to the current evidence base.

2 Literature Review

In order to illuminate the importance of early education in mathematics, this chapter begins by outlining the current political, social and educational background to the research, as well as considering what is known about typical development in mathematics and the foundational role of early skills and understanding in the preschool years. Key features of successful intervention programmes are then considered, alongside limitations of the current evidence-base. In order to provide a rationale for the present study, the potential use of educational technology as a means of raising attainment in mathematics is then explored. A systematic literature review (SLR) is nested here within the broader narrative review in order to specifically review the current evidence-base around the use of tablet-based mathematics interventions with young children. The potential interactions between young children's use of the *onebillion* intervention and other aspects of their *wider* development are then considered, including receptive language skills and positive 'approaches to learning'.¹

2.1 'Closing the gap' in mathematics achievement in the UK

Improving standards in mathematics continues to be an issue of national importance. The *Skills for Life* study (Department for Business, Innovation & Skills, 2012) indicated that approximately 24% of adults have significant difficulties with functional, everyday mathematics skills, such as paying household bills. Furthermore, the UK has performed relatively poorly in recent global comparisons, ranking only 27th out of 34 participating countries in the latest PISA² assessments of 15 year olds' mathematics attainment (Organisation for Economic Co-operation & Development, OECD, 2016). There is a particularly large gap between the attainment of the highest and lowest performing students in England (Greany, Barnes, Mostafa, Pensiero & Swensson, 2016), with a "significant tail of under achievement" that has

¹ The broader narrative review was informed by a variety of sources and papers identified through a number of searches of the literature over a two-year period, including relevant papers cited within articles identified in the SLR.

² Programme for International Student Assessment (OECD, 2016)

persisted over time (Dowker, 2009, p.4). Longitudinal research indicates that low attainment in mathematics can have significant long-term consequences, affecting later school achievement, employment, criminality, mental health and future earnings (Crawford & Cribb, 2013; Parsons & Bynner, 2006).

In many countries, underachievement in mathematics is strongly associated with social, cultural and economic disadvantage (Greany et al., 2016; OECD, 2016). Data from the PISA assessments indicate that the gap between pupils with the highest and lowest socio-economic status (SES) in England is equivalent to over three years of schooling (Greany et al., 2016; Wheeler, Durbin, McNamara & Classick, 2016). Since 2011, the UK government has introduced a number of initiatives to improve outcomes for disadvantaged pupils, such as additional 'pupil premium' funding for pupils from low income families and/or those who have been 'Looked After' by children's social care (Carpenter et al., 2013). A number of organisations have also been set up to fund and support research into educational practice, such as the Education Endowment Foundation (EEF). It is hoped that these initiatives, amongst others, will enable greater progress to be made in 'closing the gap' in mathematics achievement, particularly for disadvantaged groups.

A new primary national curriculum for mathematics was also introduced in 2014 (DfE, 2013), focusing on a 'mastery' approach to teaching mathematics, underpinned by high expectations for all and ensuring depth of understanding and fluency in mathematical content before children progress to new topics (National Centre for Excellence in the Teaching of Mathematics, NCETM, 2014). 'Mastery' learning has been adopted by other internationally high performing countries, including Singapore, Japan, South Korea and China (NCETM, 2014; OECD, 2016), and there are early indications of its success in the UK (Boylan et al., 2016). The new curriculum has, however, faced a number of criticisms for introducing demanding content too early, teaching irrelevant and outdated content, and for being informed too heavily by international practice (Alexander, 2012), at the expense of educational research (Thompson, 2012).

2.2 The importance of early years education

2.2.1 Typical development of early mathematical skills

Piaget (1941/1969) proposed that young children are not able to engage in the abstract and logical thinking required to develop a true understanding of the concept of number until they reached a stage of 'concrete-operations', at around the age of 7. At this stage of development, Piaget (1941/1969) noted that children can successfully complete 'conservation of number' tasks, showing understanding that the number of items in a set does not change when the objects are physically rearranged in space. Strong interpretations of this theory led to the view that young children may lack the cognitive capability to benefit from formal schooling in mathematics (Ginsburg & Golbeck, 2004; Hachey, 2013).

Nevertheless, more recent psychological theory and research has shown that young children demonstrate a number of mathematical competencies, even from infancy (Gelman, 2000). For example, studies have shown that 6 month old infants can discriminate between displays of dots differing in numerosity, whilst controlling for other variables, such as size and position (e.g. Xu & Spelke, 2000; Xu, 2003; Xu, Spelke, & Goddard, 2005). It is therefore thought that humans have an innate, preverbal number system for representing approximate magnitude (Feigenson, Dehaene & Spelke, 2004), which may play a role in later mathematical learning, although the mechanisms underpinning this relationship are not yet well understood (see Szkudlarek & Brannon, 2017).

As they develop, young children appear to build a more explicit understanding of mathematics through their play, interaction with others and exploration of the world (Ginsburg & Amit, 2008; Hachey, 2013). During the preschool years children develop procedures for early counting (Threlfall & Bruce, 2005), learn to solve addition/subtraction problems with small numbers (Huttenlocher, Jordan & Levine, 1994) and have been shown to engage in simple numerical reasoning, successfully predicting the outcomes of basic addition/subtraction calculations (Zur & Gelman, 2004). It is also thought that children as young as three and four can conserve number under appropriate task conditions, such as

smaller set sizes and/or reduced linguistic/attentional demands (Gelman, 1972; McGarrigle & Donaldson, 1974).

Based upon previous research, Clements and Sarama (2009) identified developmental trajectories for different areas of early mathematics, highlighting key milestones in early mathematical learning (see Table 2.1). It is, however, important to recognise all proposed age markers are approximate (Clements & Sarama, 2009), and that the level of competence children display in any particular skill may be dependent upon the learning context and the nature of the task (Ginsburg & Golbeck, 2004).

*Table 2.1: Developmental milestones in counting, number and shape
(summarised from Clements & Sarama, 2009 - see source for further details of learning trajectories in each area and other domains of mathematical development)*

Area	Key Milestones at Approximately 3 years	Key Milestones at Approximately 4 years
Counting	<ul style="list-style-type: none"> • Says number words 1-10 • Begins to develop 1:1 correspondence when counting (one number label per object), but initially there may be some errors (e.g. skipping numbers in the count sequence) 	<ul style="list-style-type: none"> • Develops an understanding that the last count label represents “how many” items there are in a set (known as <i>cardinality</i>), for small quantities and then larger amounts • Learns to count and then count out objects, initially to 10 and then later higher (to 30) • Says number words 1-20, and then later higher (to 30). • Begins to write numerals 1-10 • Begins to recognise the number one more/one less, initially by counting up in sequence • Begins to count back from 10
Number	<ul style="list-style-type: none"> • Makes a small set of items nonverbally with the same numerosity as another set which had just been shown and then hidden (NB: Children are typically able to verbally label sets of 1,2 and sometimes 3 from 1-2 years old) 	<ul style="list-style-type: none"> • Can instantly recognise small quantities (up to 4 objects) and provide the correct verbal number label (known as <i>subitising</i>)
Shape	<ul style="list-style-type: none"> • Recognises and names ‘typical’ circles, squares and sometimes triangles • Begins to match a variety of shapes with the same and then different size and orientations 	<ul style="list-style-type: none"> • Recognises circles, squares and triangles, and then rectangles, including less ‘typical’ shapes • Makes a particular ‘shape’ using concrete materials • Compares attributes of shapes, initially focusing on part and then, later, the whole shape • Identifies sides and later corners in shapes

2.2.2 Early mathematics as a foundation for later success

Longitudinal research consistently indicates that children's understanding of mathematical concepts in the early years is a powerful predictor of both mathematics and reading outcomes throughout primary and secondary education (e.g. Aubrey, Godfrey, & Dahl, 2006; Bodovski & Farkas, 2007; Duncan et al., 2007; Jordan, Kaplan, Ramineni, & Locuniak, 2009; Nguyen et al., 2016; Rittle-Johnson, Fyfe, Hofer, & Farran, 2017; Watts, Duncan, Siegler, & Davis-Kean, 2014), highlighting the importance of effective early education in mathematics to promote later learning. Ensuring that children experience early success might also promote more positive attitudes to mathematics (Krinzinger, Kaufmann, & Willmes, 2009), with additional benefits for later engagement and achievement (Chen et al., 2018; Hemmings, Grootenboer, & Kay, 2011; Mazzocco, 2009).

2.2.3 Individual differences in preschool mathematics attainment

Unfortunately, a wide gap in children's attainment is already evident before the start of formal schooling (e.g. Dowker, 2008; Howell & Kemp, 2010; Starkey, Klein, & Wakeley, 2004). Dowker (2008) noted substantial individual differences in the mathematical skills of 3-5 year old children in the UK on a range of mathematical tasks, including counting, understanding of cardinality and basic addition/subtraction skills. For example, whilst the majority of 4 year old children could count up to 10 objects correctly (70%), a substantial minority were only able to count five (16%) or three/four objects accurately (12%). National data indicate that, in 2016, approximately 20% of children did not reach the UK government 'expectations' in the areas of number, shape, space and measure by the end of the Early Years Foundation Stage (EYFS) (DfE, 2017a).

A range of factors may account for the wide variation in young children's early mathematical development. For example, research has shown that aspects of early mathematical skills are affected by individual differences observed in young children's language (LeFevre et al., 2010; Praet, Titeca, Ceulemans, & Desoete, 2013; Purpura & Ganley, 2014) and broader cognitive skills, such as attention (Dulaney, Vasilyeva, & O'Dwyer, 2015; Sims, Purpura & Lonigan, 2016). The role of language and other cognitive skills, alongside social-

emotional development, in mathematical learning is discussed further in Section 2.7.

Children's early learning experiences are also thought to have a strong influence on mathematical development, especially in the home environment (Anders et al., 2012; Melhuish et al., 2008; Napoli & Purpura, 2018; Skwarchuk, Sowinski, & LeFevre, 2014). For instance, Skwarchuk et al. (2014) found that the frequency of both formal explicit teaching of mathematics and more informal opportunities for mathematical learning at home (e.g. cooking, board games) predicted aspects of 4-5 year old children's early numeracy skills. Early years education may therefore be beneficial in ensuring that all children have access to appropriate learning opportunities, 'closing the gap' prior to the start of formal education.

Access to good quality early years educational settings is associated with better mathematical outcomes at the start of school and into both primary and secondary education (Anders et al., 2012; Melhuish et al., 2008; Taggart et al., 2015) and is often particularly beneficial for children from low SES backgrounds (OECD, 2017). A recent Ofsted report (2017), however, highlights that the quality of mathematics teaching for children in many early years classes is not as high as the teaching of literacy, and that many children are not sufficiently prepared for the demands of the new curriculum. Further guidance regarding the most effective, evidence-based instructional methods may therefore be beneficial in informing teaching practice in this area.

2.3 Early Intervention in Mathematics

This section of the review considers what *best practice* might look like in early mathematical intervention, with reference to developmental and instructional psychology, before discussing the strengths and limitations of the current evidence-base in this area.

2.3.1 What are the features of effective early intervention in mathematics?

2.3.1.1 Instructional content

There is now substantial evidence to suggest that the development of young children's understanding of number and operations, sometimes referred to as *number sense*, is an important prerequisite for future mathematical success (see Jordan, Glutting, Dyson, Hassinger-Das & Irwin, 2012). Early developing number competencies include the ability to estimate and compare numerical magnitudes, subitise (i.e. quickly recognise and label the value of small quantities), recognise numerical symbols and words, understand counting principles and perform simple addition and subtraction calculations (Howell & Kemp, 2006; Jordan, Kaplan, Nabors Oláh, & Locuniak, 2006; Jordan et al., 2012). As predicted by theories of cumulative learning (Gagné, 1968), mastery of these basic concepts and skills may be foundational for later learning, predicting success with more complex mathematical skills (e.g. Nguyen et al., 2016; Rittle-Johnson et al., 2017). Longitudinal studies indicate that early number sense at 5-6 years is also predictive of the *rate* of progress that children make in mathematics throughout later primary education (Jordan et al., 2006; Jordan et al., 2009; Marcelino, de Sousa, Cruz, & Lopes, 2012)

Despite the dominant focus on number in the research literature, the importance of developing young children's non-numerical competencies should also be acknowledged. For example, it is thought that children's pattern recognition skills may support analogical reasoning, enabling children to identify rules in the number system (Rittle-Johnson et al., 2017). In a recent longitudinal study in the United States (US), Rittle-Johnson et al. (2017) showed that the patterning skills of 4-5 year old children predicted later mathematical skills at the age of 10-11. Similarly, Nguyen et al. (2016) found that early patterning, geometry and measurement skills predicted later mathematical skills at 10-11 years, although number skills, and particularly advanced counting, cardinality and subitising skills, were the strongest predictors of later achievement. Further research is, however, needed to replicate these findings, particularly in the UK educational system, and to fully understand the mechanisms by which non-numerical skills may support overall mathematical development.

2.3.1.2 *Instructional approaches*

Consideration must also be given to the most effective, evidence-based methods for delivering curriculum content. Empirical research has, for example, frequently shown that systematic, explicit instruction in mathematics is highly effective for both young children and those at-risk of mathematical difficulties (Baker, Gersten, & Lee, 2002; Carbonneau & Marley, 2015; Doabler, Fien, Nelson-Walker, & Baker, 2012; Gersten & Carnine, 1984; Gersten et al., 2009). Explicit, or direct, instruction in mathematics typically involves the following core instructional elements: clear models for learning, opportunities for guided practice and regular academic feedback, through a carefully staged curriculum (Doabler & Fien, 2013). Compared to discovery-based learning approaches, children are given high levels of instructional guidance to ensure that they achieve success (Mayer, 2004) and to reduce demands on working memory (Kirschner, Sweller, & Clark, 2006). In a review of 11 studies examining interventions for students with low-mathematics attainment, Gersten et al. (2009) found that explicit instruction had a large positive benefit for attainment ($d = +1.22$). Targeted, direct instruction may therefore be important for young children, alongside more informal, play-based opportunities for learning mathematics.

Personalisation in the delivery of the curriculum may also be essential. As well as individual differences in young children's overall mathematical skills, it is well-documented that there is wide variation in children's understanding of different aspects of mathematics (Dowker, 2005). Instructional programmes should therefore acknowledge the strengths and needs of each individual, enabling children to build upon their previous knowledge and understanding (Gifford & Rockliffe, 2012; Holmes & Dowker, 2013).

Children's mathematical learning may also be supported by access to visual models to support their understanding of more abstract concepts, such as number lines, hundred squares and finger models (Doabler et al., 2012; Gersten et al., 2009). In the early stages of skill acquisition, however, use of concrete manipulatives, such as counting blocks and 3D shapes, are often recommended (Doabler et al., 2012). The value of moving from concrete to pictorial and then to more abstract forms of representation (the CPA approach;

Gifford, Black & Griffiths, 2015), is underpinned by Bruner's (1966) enactive-iconic-symbol theory of mathematical development. Bruner proposed that young children initially represent mathematical knowledge in terms of motor action (enactive phase). Once this is consolidated they can then move to more abstract understanding, representing their knowledge in mental images (iconic phase), until, finally, they are able to use written symbols and words (symbolic phase). Nevertheless, research indicates that the effectiveness of manipulatives depends upon how they are implemented in the classroom: children benefit from familiarity with the materials and teacher guidance in how to use them appropriately (Carbonneau & Marley, 2015; Carbonneau, Marley, & Selig, 2013; Gifford et al., 2015).

2.3.2 How effective are current early years mathematics interventions?

A growing number of studies, predominantly conducted in the US, have evaluated the impact of early mathematics interventions for young children. Interventions have been implemented at a universal level, as well as a more targeted level for children who may be at-risk of mathematical difficulties (Wang, Firmender, Power, & Byrnes, 2016). In a recent meta-analysis, Wang et al. (2016) demonstrated that, on average, early intervention programmes have a moderate to large benefit on the mathematical attainment of children under six. A strength of many of the current intervention programmes is their use of theoretically-informed instructional approaches, as discussed above. The *Roots* programme, for example, aims to develop children's proficiency in counting, numerical comparison, simple calculation and place value (Clarke et al., 2016). Children receive explicit instruction, including careful teacher modelling, guided practice and immediate, specific academic feedback. Visual representations (e.g. number lines) and concrete manipulatives (e.g. counting and place-value blocks) are also incorporated into lessons. A recent evaluation found that the intervention had positive short-term benefits for 5-6 year old children (Clarke et al., 2016).

2.3.2.1 Challenges in early intervention research

A number of challenges, however, remain in this field of research. First, a consistent finding is that the effects of interventions diminish over time; children

who have not received the intervention typically catch-up with others who have (Bailey et al., 2016). Longitudinal fade-out, for example, was found following the *Roots* intervention (Clarke et al., 2016). Such findings perhaps run contrary to cumulative learning theories (Gagné, 1968) that predict earlier learning will lay the foundations for future skill building, as well as correlational evidence, as noted above, which shows strong associations between early mathematics and later achievement.

A number of alternative explanations might account for fade-out effects. First, it may be that new knowledge that children have acquired may not be fully consolidated, leaving it more liable to interference and forgetting (Bailey et al., 2016; Wixted, 2004). This explanation is consistent with predictions from the Instructional Hierarchy (Alberto & Troutman, 1986, Haring & Eaton, 1978; Martens & Witt, 2004) (see Figure 2.1), which suggests that learning skills to a high level of fluency (both with accuracy and *speed*) and opportunities for 'overlearning' (continued practice after a skill is first acquired) may contribute to maintenance over time. Alternatively, early interventions may not influence relatively stable factors (such as environmental influences, motivation or cognitive skills, e.g. executive functioning or language), which continue to impact upon children's mathematical learning as they grow older. This hypothesis has received tentative support from analysis of fadeout effects from one programme, *Building Blocks* (Bailey et al., 2016). Consideration should also, however, be given to the curriculum that children receive post-intervention and the extent to which it builds upon the new skills that children have acquired (Bailey et al., 2016); some studies have shown, for instance, that there are greater longitudinal benefits when follow-through intervention is provided in subsequent year groups (Clements, Sarama, Wolfe, & Spitler, 2013). It is likely that a combination of these factors affect the longitudinal success of interventions, with differing impacts depending upon the nature of the intervention concerned.

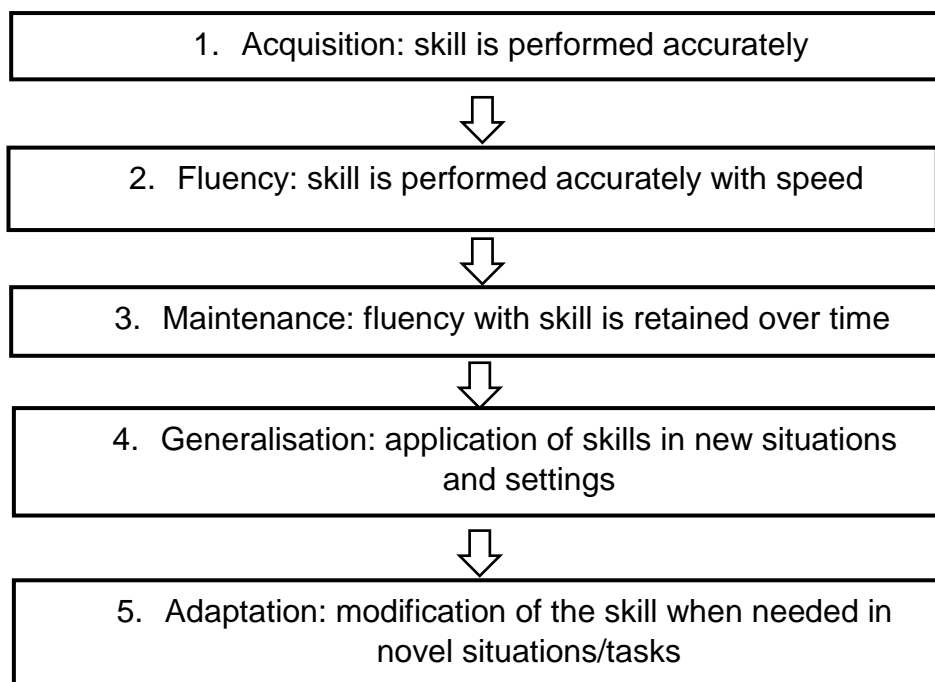


Figure 2.1: The Instructional Hierarchy (adapted from Alberto & Troutman, 1986; Haring & Eaton, 1978; Martens & Witt, 2004)

An additional challenge for researchers is that significant effects are reported more frequently on researcher-designed measures rather than standardised assessments (Gersten, 2016; Wang et al., 2016). Researcher-designed assessments are typically *proximal* measures of attainment, testing skills closely aligned to the content of the intervention, and are therefore sensitive to learning gains, although they may also be more subject to researcher bias and lower reliability. In contrast, standardised assessments are often more *distal* to the taught content, and may, in part, assess whether learners are able to transfer and apply their learning to new contexts that have not been directly encountered during the intervention (Gersten, 2016), consistent with the generalisation and adaptation stages of the Instructional Hierarchy (Haring & Eaton, 1978), see Figure 2.1.

It is also important to note that not all children may benefit equally from an intervention. Individual differences in response to intervention (RtI) are frequently reported across early mathematics intervention studies (Fuchs,

Fuchs & Compton, 2012; Salminen, Koponen, Leskinen, Poikkeus & Aro., 2015). Understanding the causes of these differences and ensuring sufficient personalisation in the delivery of the curriculum may therefore be essential in supporting positive outcomes for all (Gifford & Rockliffe, 2012; Holmes & Dowker, 2013).

Whilst showing promise, further work is therefore needed to ensure that early mathematics interventions can lead to sustained and generalised mathematical learning.

2.4 The Role of Technology in Early Mathematics Intervention

One possible approach to early intervention in mathematics may be through the use of educational technology, which can offer personalised programmes of instruction to young learners in accessible and engaging formats. Since the 1980s, schools have been experimenting with different types of technology to support learning, including: computers, interactive whiteboards and, increasingly, mobile devices, such as tablets (Cheung & Slavin, 2013). The use of educational technology in schools and other educational settings continues to rise across the globe (Hubber et al., 2016). Nevertheless, Ofsted (2012) highlights that there is “underdeveloped use of information and communication technology to enhance learning” (p.20) in schools across the UK.

Although the term educational technology has been defined in various ways in the literature, the present review will adopt the following working definition, developed by Cheung and Slavin (2013):

“Educational technology refers to a variety of technology-based programs or applications that help deliver learning materials and support the learning process...to improve academic learning goals (as opposed to learning to use the technology itself).” (p. 90)

A rationale is presented here for how technology may enhance outcomes within early years’ mathematics education, whilst considering some of the concerns and challenges raised for its use with young children and implementation in the classroom.

2.4.1 How can the use of tablet technology benefit young children's learning in mathematics?

Whilst a number of computer-based mathematics interventions have been developed for young children (e.g. Fien et al., 2016; Praet & Desoete, 2014), a range of programmes are now available for use on tablets. The popularity of tablets for young children is increasing. Surveys suggest that approximately 55% of 3-4 year old children use a tablet at home (Ofcom, 2016). For young children, tablets may be particularly beneficial as (a) they remove the need for additional devices such as a mouse and keyboard that require manual dexterity, (b) they are portable and lightweight and (c) they allow access to a range of software, known as applications, or 'apps', specifically designed for children in the first few years of school (Kucirkova, 2014; Neumann & Neumann, 2014). Research indicates that the majority of 3-4 year old children are able to utilise many of the key features of tablet technology independently, such as opening apps, swiping the screen and tracing shapes (Marsh et al., 2015).

As discussed in Section 2.5.1.2, well-designed app software has the potential to incorporate a wide range of evidence-based instructional approaches into the context of play-based learning (Outhwaite et al., in press). Similarly to traditional play, computer play can lead to increases in children's self-motivation, experimentation and higher levels of engagement (Verenikina, Herrington, Peterson & Mantei, 2010).

2.4.2 What impact does educational technology have on mathematical attainment?

A number of meta-analyses have been conducted to evaluate the impact of technology based programmes on children's mathematical attainment, as shown in Table 2.2. They consistently show that educational technology can have positive benefits for learning outcomes, although there is considerable variation in the effect sizes³ (ESs) reported across reviews. Variations might reflect the timescale of the review, the criteria selected for study inclusion and the analysis methods selected. Two meta-analyses have shown that younger

³ Effect sizes, standardised measures of the magnitude of intervention effects, are discussed further in Section 4.2.1 in relation to analysis procedures for the present study.

children tend to benefit more from technology-based interventions than older students (Cheung & Slavin, 2013; Li & Ma, 2010), although none of the reviews examine the impact of technology for children under the age of five.

Table 2.2: Effect sizes from meta-analyses evaluating the impact of educational technology on children's mathematics attainment

Meta-analysis	Age Range	Years Reviewed	Effect Size
Chauhan (2017)	5-10 years	2000 - 2016	+0.47
Cheung and Slavin (2013)	5-18 years	1980 - 2010	+0.16
	5-10 years	1980 - 2010	+0.17
Li and Ma (2010)	5-18 years	1990 - 2006	+0.28
	5-10 years	1990 - 2006	+0.78
Slavin and Lake (2008)	5-10 years	1971 - 2006	+0.19

The most recent meta-analysis conducted by Chauhan (2017) indicates that educational technology has a moderate positive impact upon attainment in mathematics. This review may be of greatest relevance in assessing the impact of currently available technology, such as tablet interventions, although due to the date of publication, a number of more recently published studies from April 2016 onwards are not included (e.g. Outhwaite et al., 2017, Schacter & Jo, 2017), as discussed further below.

Positive outcomes have also been reported across a number of studies exploring the impact of tablet technology for learning more generally (Haßler, Major & Hennessey al., 2016; Lovato & Waxman, 2016). Nevertheless, the current evidence base is relatively limited, with reviewers calling for greater experimental rigour, such as longitudinal evaluations of tablet technology over time (Haßler et al., 2016; Kucirkova, 2014). It is also important to note that the efficacy of the intervention may be highly dependent on the nature of the software itself. Many apps, for example, may focus on memorisation, rather than developing conceptual understanding (Larkin, 2013). The pace of

technological advancement is currently outstripping research in this area, with further research required to ensure that educational practice is evidence-based.

2.4.3 Criticisms regarding the use of educational technology with young children

The use of technology by young children remains somewhat controversial. Concerns have been voiced by children's advocates and the research community that high levels of 'screen-time' have negative consequences for children's social-emotional development, obesity levels, language and cognitive development (Cordes & Miller, 2000; Sigman, 2012). The American Academy of Pediatrics (AAP) has recently recommended that screen-time is limited to 1 hour per day for 2-5 year old children, and restricted to the use of "high quality" educational programmes (AAP, 2016, p.3), although further evidence is required to support these recommendations.

The effects of screen-time on overall development may depend, though, upon the nature of technology use, such as the type of media, activity and whether it is used in interaction with others (Herodotou, 2018; Lovato & Waxman, 2016). It is important to recognise that the wider developmental impact of tablet use is not well understood at present (Herodotou, 2018; Lovato & Waxman, 2016) and warrants further research, particularly concerning which programmes can be deemed 'high quality' and the conditions under which tablet use and educational technology is most beneficial (Higgins, Xiao & Katsipataki, 2012). There is growing recognition that implementation processes (e.g. participant characteristics and environmental context) often have a significant influence on the outcome of interventions (Cook & Odom, 2013; Horner, Sugai, & Fixsen, 2017; Nordstrum, LeMahieu, & Berrena, 2017). Aligning with the 'implementation science' movement, an understanding of "what works, for whom, in what circumstances" (Pawson & Tilley, 1997, p.84), may be imperative in guiding best practice in the use of educational technology with young children.

2.4.4 What aspects of implementation might affect learning outcomes from educational technology?

Numerous aspects of implementation might affect learning outcomes from tablet programmes, such as the duration and frequency of the intervention, the classroom environment and how the use of technology is embedded within the wider curriculum. One particularly important factor, however, may be how children are supported in their use of tablet technology by teaching staff. There is, for example, preliminary evidence that outcomes from tablet apps are enhanced when they are used in a social context (Teepe, Molenaar & Verhoeven, 2016; Walter-laager et al., 2016). The role of facilitator support is therefore given additional consideration here.

2.4.4.1 Role of the facilitator

Vygotsky (1978) proposed that with additional social support, or scaffolding (Wood, Bruner & Ross, 1976), children are able to achieve a higher level of competence in tasks. The gap between what a learner can do with and without scaffolding is referred to as the *zone of proximal development (ZPD)* (Vygotsky, 1978). In the context of instruction with educational technology, Yelland and Masters (2007) propose that three particular forms of teacher scaffolding might enhance children's learning: cognitive, technical and affective. *Cognitive scaffolding* involves the teacher providing additional questioning, modelling and assistance to the child so that they can master the particular skill or concept presented on the device. This support may be more targeted to the needs of the individual than software-based scaffolding. Facilitators might also provide *technical scaffolding*, maximising children's use of the in-built features of the software, which may in turn, benefit their understanding. Finally, children may need some *affective scaffolding* in the form of praise and encouragement to stay on task and to progress to higher levels of understanding.

In broadly qualitative case studies with 7-8 year old children, Yelland and Masters (2007) demonstrate how all three types of scaffolding can be beneficial whilst children are engaged in computer-based tasks. Some quantitative data were also presented, indicating that learning outcomes appeared to be greater in a scaffolded environment; however, due to the small-scale of the research,

inferential statistical analysis was not used, limiting the conclusions that can be drawn. Other qualitative research has also emphasised the benefits of adult support and scaffolding, particularly when young children encounter technical difficulties with software (Couse & Chen, 2010; Matthews & Seow, 2007). Nevertheless, further research is required to understand the most appropriate facilitator approaches when support learning through technology.

2.5 The *Onebillion* Tablet-Based Mathematics Intervention

Overall, the current literature indicates that educational technology may have potential benefits for the mathematics development of young learners, but further research is required to guide educators, EPs and policy-makers about the most effective software and forms of implementation. The present research therefore aims to contribute to the current evidence-base by evaluating a mathematics intervention developed by the charity *onebillion*, consisting of two tablet apps, *Maths Age 3-5* (Version 1.4, *onebillion*, 2016) and *Maths Age 4-6* (Version 1.2, *onebillion*, 2016)⁴. Given the similarity of instructional features across both apps, they are treated as one continuous intervention programme for research purposes.

2.5.1 What features of the *onebillion* software may support mathematical development?

2.5.1.1 Instructional content

The instructional content of the apps follows the UK National Curriculum (DfE, 2017b) for this age range; children work through a series of activities organised into different topic areas, including: sorting and matching, counting, prepositional language, patterns and shapes, size/quantity comparisons and adding/taking away. There is a particular instructional focus on understanding of numbers to 20, counting skills and basic calculation, all of which aim to develop children's number sense, an important contributor to later learning (see Section 2.3.1). However, the apps also provide instruction in other non-numerical

⁴ The *onebillion* tablet apps, *Maths Age 3-5* and *Maths Age 4-6*, are copyrighted and paid-for applications. Further information about the content of the apps and purchasing the intervention is available from the *onebillion* (2018) website (<https://onebillion.org/apps/>, accessed 24.07.18).

competencies within the shape, space and measure aspects of the UK curriculum, including patterning.

2.5.1.2 Instructional processes

The apps incorporate a number of the principles of evidence-based instructional practice in mathematics (see Section 2.3.1). A ‘virtual’ teacher delivers the ‘lessons’ through explicit teaching methods providing clear modelling and instructions for each topic area, followed by opportunities for independent practice and academic feedback; children receive ‘stars’ and ‘ticks’ for correct responses and an error noise following a mistake. Feedback is instantaneous, potentially promoting greater engagement and progress compared to traditional teaching methods, as children do not have to wait for attention from an adult (Haake, Husain, & Gulz, 2015; Henderson & Yeow, 2012). Moreover, the simultaneous presentation of auditory and visual input supports multi-sensory learning of mathematical concepts, shown to enhance understanding (Carr, 2012; Pavio, 1986).

Each child also has a unique software profile and is able to regulate their own pace of learning through a staged, task-sliced curriculum to support mastery (Magliaro, Lockee, & Burton, 2005), potentially helping to match instruction with the child’s ZPD (Vygotsky, 1978), reducing frustration (Haake et al., 2015) and creating a personalised learning environment. Children complete a quiz at the end of each topic, progressing forward once they achieve a 100% pass rate in order to ensure that they have achieved mastery in achieving a specific learning goal; they can return to earlier activities for further practice if needed, potentially increasing the fluency of skill learning, and supporting maintenance/generalisation (see Section 2.3.2, regarding the Instructional Hierarchy, Alberto & Troutman, 1986; Haring & Eaton, 1978; Martens & Witt, 2004). Children receive a virtual certificate after passing an app quiz, providing additional positive reinforcement to support learning.

Consistent with the CPA approach (Gifford et al., 2015, see Section 2.3.1), the *onebillion* app activities contain a number of clear visual representations of mathematical concepts to support mathematical learning. Moreover, just as traditional mathematical learning in the early years often uses concrete

manipulatives, the *onebillion* software involves children accessing virtual manipulatives, where an interactive object can be manipulated on screen (Moyer-Packenham et al., 2015). For example, children can move familiar objects to support matching and understanding of number. Research indicates that virtual manipulatives can be just as effective as traditional manipulatives in enhancing mathematical outcomes for young children (Mattoon, Bates, Shifflet, Latham, & Ennis, 2015).

Use of the apps is typically facilitated by an adult, who can provide additional cognitive, affective and technical scaffolding where necessary (Yelland & Masters, 2007).

2.5.2 How effective is the *onebillion* tablet intervention?

There have been a number of previous evaluations of the *onebillion* intervention in the UK (Outhwaite et al., in press; Outhwaite et al., 2017) and related apps in Malawi (Pitchford, 2015; Pitchford, Kamchedzera, Hubber & Chigeda, 2018), indicating that the apps hold promise for early years mathematics education. However, in order to inform the nature of the present research, a systemic review was conducted to objectively evaluate the current evidence-base around the use of the *onebillion* intervention and related tablet-based mathematics interventions for young children.

2.6 Systematic Review of the Evidence-Base

Systematic literature reviews (SLRs) aim to synthesise findings across the literature in order for the researcher to establish what is known about a particular field, critique the current evidence base and identify gaps for further research (Andrews, 2005). Systematic reviews are therefore ideally placed to address questions about ‘What works?’ and the effectiveness of educational interventions (Gough, Oliver, & Thomas, 2012; Petticrew & Roberts, 2006).

The purpose of the SLR conducted here was to address the following question:

What is the impact of tablet-based interventions on young children’s mathematics attainment?

2.6.1 Systematic review methodology

Unlike other types of review, the term *systematic* emphasises the need to ensure that the methods used in searching, selecting and appraising studies are explicitly stated, ensuring that the findings can be replicated and that bias is minimised (Gough et al., 2012). The review follows the key methodological steps suggested by Gough (2007), including: defining eligibility criteria, searching and screening studies, describing studies, appraising their quality and relevance and then presenting a synthesis of findings.

2.6.2 Eligibility criteria

Eligibility criteria were established to determine the studies that would be included within the review (Table 2.3).

Table 2.3: Eligibility criteria for studies to be included in the systematic review

Study Feature	Eligibility Criteria
Design	<ul style="list-style-type: none">• Pre- and post-test control or comparison group (technology intervention) designs (including quasi-experimental or randomised control trials (RCT)s)
Participants	<ul style="list-style-type: none">• At least some participants aged 6 years and under• Participants not restricted to a population with specific difficulties (e.g. hearing impairment)
Intervention	<ul style="list-style-type: none">• A mathematics-specific intervention delivered directly through tablet technology• Delivered to children over more than one session• A clearly defined tablet programme - <i>studies were excluded if they involved access to a number of intervention programmes with different instructional features or only used the technology as a tool to support the wider lesson.</i>
Outcome Measures	<ul style="list-style-type: none">• Quantitative outcomes on early mathematics understanding and skills
Context	<ul style="list-style-type: none">• Study took place outside of a clinic or laboratory setting (e.g. child's home, nursery or school)
Additional Features	<ul style="list-style-type: none">• Research published in a peer-reviewed academic journal• Research published in the English language

2.6.3 Database search

A number of databases (Web of Science, Eric (via EBSCO) and PsycINFO) were then searched, identifying articles containing the following terms in the abstract, title or key words of the article, as shown in Table 2.4.

Table 2.4: Search terms included in the systematic review

Main Search Term	Alternatives	Truncations or Alternative spellings
Tablet	iPad ⁵ , handheld, mobile	tablet*, iPad*, mobile* hand-held
Mathematics	numeracy, arithmetic	math*
Intervention	programme, game	program*, game*
Educational (purpose/context)	teaching, learning, school, classroom, home, parent	home*, parent*, school*, teach* education*, class*, learn*
Early Years (age range)	preschool, primary, kindergarten, nursery, Grade 1, elementary, Foundation Stage, Year 1, Year 2	pre-school, first grade

The search was limited to studies published in the last 15 years, 2003-2018, due to advancements in educational technology. Searches were also limited to peer reviewed, published journals, available in English.

The search yielded the results displayed in Table 2.5.

Table 2.5: Database search results (search conducted 06/04/18)

Database	Number of Records
Web of Science	49
ERIC (via EBSCO)	118
PsycINFO	23

An additional article was identified by the author (Outhwaite et al., in press) which reports on findings from a randomised control trial (RCT) evaluation of

⁵ iPad is a trademark of Apple Inc., registered in the US and other countries. This thesis is an independent publication and has not been authorised, sponsored, or otherwise approved by Apple Inc.

the efficacy of the *onebillion* mathematics intervention. Although still in press, this article was also included in the synthesis due to the direct relevance of the findings for the present study.

2.6.4 Screening

Any duplicate articles were then removed and the title/abstracts of the remaining studies were screened. Full-text articles were then assessed against the eligibility criteria before inclusion into the final synthesis, as outlined in Figure 2.2.

Screening resulted in the identification of 11 research articles. One article (Outhwaite et al., 2017) reported on results from four studies, two of which met eligibility criteria (referred to by the authors as Study 1/Study 4), and therefore 12 unique studies were included in the final synthesis.

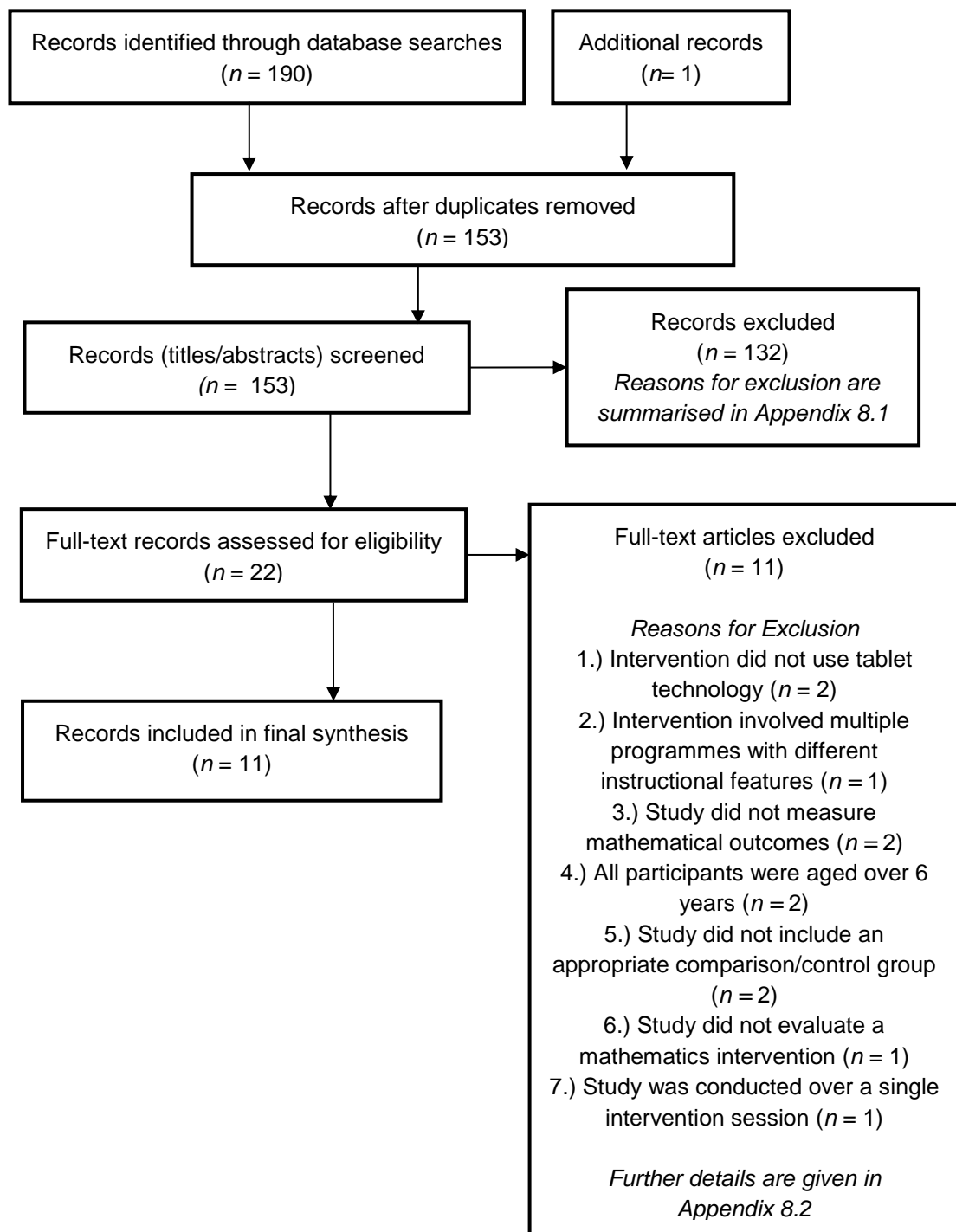


Figure 2.2: The process of screening and selecting records for the systematic review (adapted from Moher, Liberati, Tetzlaff, Atman & the PRISMA Group, 2009)

2.6.5 Quality and relevance of research

Before synthesising the research evidence across studies in a SLR, it is important to appraise the overall quality of each study and the relevance of its findings in addressing the review question (Gough, 2007). These judgements enable the reviewer to determine the relative weighting that this study should be given in the overall synthesis of research evidence (Andrews, 2005).

Gough (2007)'s Weight of Evidence (WoE) framework outlines three dimensions against which the quality of research can be appraised: (A) the quality of the research design; (B) the appropriateness of the research design to answer the review question; and (C) the relevance of the findings in answering the review question. The WoE framework does not specifically define the criteria against which judgements should be made within these three areas, as these will depend upon the specific aims and focus of the review (Gough, 2007). The flexibility afforded by this framework and its ability to evaluate studies employing a range of different designs make it appropriate for reviews focused on the evaluation of 'real-world' educational research. It was therefore considered to be a suitable method for assessing study quality and relevance in the present review.

Each study was rated as 'high', 'medium' or 'low' against the three WoE dimensions and ratings were averaged to give an overall WoE judgement (D) for each study. Study ratings and the specific criteria used to make these judgements are detailed in Appendix 8.3. The possibility of bias is acknowledged given that the criteria were developed and judged only by the author.

2.6.6 Critical analysis of the evidence-base

The evidence base was then synthesised and critically analysed to determine the extent to which it is able to address the research question and to identify potential implications for future research. Core themes for analysis were drawn from the WoE model, including: sample characteristics, design/analysis, intervention, outcome variables and measures. Key characteristics of each study are detailed within Appendix 8.4 to support comparison.

2.6.6.1 Sample characteristics

Of the twelve studies included in this review, the majority have been delivered at a universal level; only one study examined the impact of a tablet-based intervention as a targeted programme for low-attaining children (Outhwaite et al., 2017-Study 4). Five studies, however, conducted a separate analysis of findings for children identified as having lower attainment at pre-test (Hieftje, Pendergrass, Kyriakides, Gilliam, & Fiellin, 2017; Outhwaite et al., in press; Schacter & Jo, 2016, 2017; Schacter et al., 2016). A number of researchers have also evaluated the efficacy of technology based intervention for children in disadvantaged groups, focusing upon samples from low socio-economic backgrounds (Outhwaite et al., 2017-Study 1; Park, Bermudez, Roberts, & Brannon, 2016; Schacter & Jo, 2016; Schacter et al., 2016) or developing countries (Pitchford, 2015).

Table 2.6 displays the age distribution of the participants receiving the interventions across each study. In seven of the studies at least some children receiving the intervention were over the age of six; consequently, it is difficult to determine the extent to which reported effects might be different if only younger children were included. Only two studies included 3-4 year old children in the equivalent school year to Foundation Stage 1 in the UK (Kosko & Ferdig, 2016; Park et al., 2016), indicating a weak evidence base at present for the use of technology-based interventions with nursery aged children.

Table 2.6: Age distribution of participants across studies included in the systematic review

Study	Age (years)					
	3	4	5	6	7	8+
Berkowitz et al. (2015)						
Hietje et al. (2017)						
Kosko and Ferdig (2016)						
Outhwaite et al. (2017) Study 1						
Outhwaite et al. (2017) Study 4						
Outhwaite et al. (in press)						
Park et al. (2016)						
Pitchford (2015)						
Schacter and Jo (2016)						
Schacter et al. (2016)						
Schacter and Jo (2017)						
van der Ven, Segers, Takashima, & Verhoeven (2017)						

Sample sizes vary considerably across studies, from 27 through to 587; five studies included a sample of $n < 100$ (see Appendix 8.4). A number of studies also only conducted research with children from one educational setting (Outhwaite et al., 2017-Studies 1/4; Pitchford, 2015; Schacter et al., 2016). Limitations in sample size might limit the external validity of the findings. Furthermore, only three studies have been conducted in the UK (Outhwaite et al., in press, 2017-Studies 1/4). Due to wide variations in the curriculum, pedagogical approaches and socio-cultural factors between countries, further research is needed to test the efficacy of technology-based interventions in the UK education system.

2.6.6.2 Design and analysis

All but one of the studies included in the review were given a ‘high’ to ‘medium’ rating for the overall quality of the research design (WoE A rating). A strength of the current evidence base is that the majority of studies (9 out of 12) employed a RCT design, where participants or schools were randomly allocated to control or experimental groups (see Appendix 8.3). This design is thought to have high internal validity and is often referred to as the “gold standard” of evaluation

research (Cook, 2009, p.1). The remaining seven studies used quasi-experimental designs, which have lower internal validity due to the risk of selection bias. In two of these studies, Outhwaite et al. (2017) Study 1 and 4, groups were not equivalent on pre-test measures and these differences were not controlled through statistical analysis, weakening their WoE B judgement.

The nature of control and/or comparison groups varied across the reviewed studies, potentially affecting the comparability of the findings. Seven studies used a 'treatment as usual' (TAU) control group, where children continued with their normal educational programmes (see Appendix 8.3). In three of these studies, a control group received mathematics instruction whilst the experimental group accessed the intervention (Outhwaite et al., in press, 2017-Study 4; Pitchford, 2015). This design has the advantage of approximately equating mathematical instruction time across groups, but may raise some ethical concerns due to the unknown benefits of a new intervention as compared with typical instruction.

Six studies included a comparison group accessing educational technology within the research design (see Appendix 8.3). Inclusion of a comparison group enables an assessment of whether improvements are due to the content of the intervention itself, rather than improvements in children's motivation, manual skills or attention that might result from using technology in general. It could, however, be argued that the hardware is part of the overall 'package' of a complex intervention (Cheung & Slavin, 2013), and the potential benefits of the programme might be underestimated unless a TAU control group is also incorporated within the design (e.g. Pitchford, 2015). This may be a particularly relevant consideration in interpreting studies conducted by Schacter et al. (2016) and Schacter and Jo (2017), as the comparison group used other maths-based tablet apps.

Only two studies incorporated a longitudinal follow-up in the research design (Outhwaite et al., 2017-Study 1; van der Ven et al., 2017). Further research involving longitudinal designs is therefore needed to determine whether early interventions have a sustained impact on children's mathematics attainment.

2.6.6.3 *Intervention*

Specific features of the tablet-based interventions are contrasted in Table 2.7. The majority of interventions were delivered in educational settings; in three studies, however, they were implemented in children's homes or in after-school programmes. Although all of the studies identified through screening have relevance in answering the review question, higher WoE C ratings were given to studies conducted in educational settings as the present research was implemented in a nursery context.

Interventions contained a wide variety of different features which might affect their efficacy (see Table 2.7) although commonly occurring features include staged progression of activities, immediate feedback, virtual manipulatives and rewards for positive reinforcement. Two studies included interventions focused specifically on developing addition and subtraction skills (Park et al., 2016; van der Van et al., 2017).

Regarding implementation, the majority of studies involved a member of teaching staff or a parent facilitating the child's use of the intervention, see Table 2.7. Schacter et al. (2016) used a researcher facilitator, potentially reducing the ecological validity of the research. Across the published studies, few details are given about the nature/level of support that facilitators provided to children and qualitative data is not provided to illuminate how aspects of implementation may have affected outcomes.

It should be noted that the duration of the interventions varied significantly across studies included in the review, from 2-3 weeks (Park et al., 2016; Kosko & Ferdig, 2016) to a full school year (Berkowitz et al., 2015), see Appendix 8.4. It may be difficult to draw conclusions about the full impact of interventions implemented over only a short time scale, particularly where the effects of novelty on children's attention and motivation might decline over time.

Table 2.7: Key features of interventions included in the systematic review

Name of Intervention	Study or Studies	Setting	Implementation Context	Intervention Content	Distinguishing Software Features (summarised from study article/s)
Bedtime Learning Together (Developed by: <i>Bedtime Math Foundation</i>)	Berkowitz et al. (2015)	Home	Parent and child using the app together and discussing the content	Topics include geometry, arithmetic, fractions, counting and probability	<ul style="list-style-type: none"> • Story passages about a particular maths topic with five corresponding questions to then answer
Knowledge Battle (Developed by: <i>Yogome, Inc.</i>)	Hieftje et al. (2017)	After-school programmes in schools/ community settings	Not discussed	Tasks to develop mathematics skills in the Mexican curriculum, consisting of 21 mini-games	<ul style="list-style-type: none"> • 21 mini-games, increasing in difficulty, inside a larger battle-style game • Characters/story-based learning • Mastery and repetition of concepts • Reward (earning <i>Power Cubes</i>) • Immediate feedback and explanation of incorrect responses
Zorbit (Developed by: <i>Clockwork Fox Studios</i>)	Kosko and Ferdig (2016)	Home	Children received varying levels of support from their parents	Tasks to develop number recognition, sorting/matching, counting, quantity comparison, understanding of ordinal numbers, spatial reasoning and geometry	<ul style="list-style-type: none"> • Six levels of progressive difficulty • Story-based learning • Reward/feedback included (stars to build a rocket ship)

Name of Intervention	Study or Studies	Setting	Implementation Context	Intervention Content	Distinguishing Software Features (summarised from study article/s)
Two continuous apps: Maths Age 3-5 and Maths Age 4-6 (Developed by: <i>onebillion</i>)	Outhwaite et al. (2017) Study 1 and 4 Outhwaite et al. (in press)	School	Children used the intervention in small groups in their typical classroom with the support of a member of teaching staff	Tasks to develop number, shape, space and measure concepts in the UK National Curriculum	<ul style="list-style-type: none"> • Virtual teacher provides modelling • Immediate feedback • Progressive levels of difficulty • Virtual manipulatives • Reward (certificates/stars) • Assessment quizzes to ensure mastery of taught content
Approximate arithmetic training (Developed by: <i>M. Paris, known to study authors</i>)	Park et al. (2016)	School	Children used the app in small groups. The nature/role of the facilitator was not specified.	Tasks focus upon addition and subtraction of arrays of objects	<ul style="list-style-type: none"> • Characters to introduce activities • Immediate feedback • Difficulty manipulated based on past performance • Cartoon videos shown at regular intervals to increase motivation
Four related apps: Masumu 1, Masumu 2, Count to 10 and Count to 20 (Developed by: <i>onebillion</i>)	Pitchford (2015)	School	Children used the apps in small groups in a separate classroom with the support of teaching staff	Tasks to develop basic number understanding, following the National Primary Curriculum in Malawi	<ul style="list-style-type: none"> • Virtual teacher provides modelling • Immediate feedback • Progressive levels of difficulty • Virtual manipulatives • Reward (stars) • Assessment quizzes to ensure mastery of taught content

Name of Intervention	Study or Studies	Setting	Implementation Context	Intervention Content	Distinguishing Software Features (summarised from study article/s)
Math Shelf (Developed by: <i>J. Schacter</i>)	Schacter and Jo (2016) Schacter and Jo (2017) Schacter et al. (2016)	School	Children used the intervention in their typical classroom with the support of a member of teaching staff for the first two weeks, and then independently for the remaining weeks Children used the intervention in a small group in a separate classroom, supervised by a member of the research team	Tasks to develop subitizing, counting, comparison, sequencing, number recognition and place value	<ul style="list-style-type: none"> • Virtual manipulatives • Progression of activities based upon Montessori sequence • Placement test to determine where students begin • Modelling of activities • Immediate feedback • Scaffolding (cues to support learning and additional support if children are first unsuccessful) • Reinforcement • Monitoring and repetition of activities where children needed additional scaffolding
Racing game (Developed by: <i>study authors in cooperation with QLVR software designers</i>)	van der Ven et al. (2017)	School	Children used the intervention supervised by a member of teaching staff	Tasks to develop arithmetic fluency with addition/subtraction	<ul style="list-style-type: none"> • Progressive levels of difficulty • Greater reward from accuracy and increased speed in responding • Immediate feedback • Competition-element (race between two cars) and opportunity to earn points to upgrade car/scenery

2.6.6.4 Mathematics measures and dependent variables

A wide variety of measures were used to assess mathematical outcomes across the studies included in the review (Appendix 8.4). The majority of studies assessed the impact of the intervention on children's early number skills and conceptual understanding. Only six studies explored the impact of educational technology beyond the domain of number to look at broader mathematical knowledge and skills, such as shape, pattern, sorting/matching and measurement (Hieftje et al., 2017; Kosko & Ferdig, 2016; Outhwaite et al., in press, 2017-Studies 1/4; Pitchford, 2015).

A limitation of the evidence base is that the majority of studies did not use standardised measures (Appendix 8.3), with researchers perhaps seeking to develop measures that might be more sensitive to experimental gains. Inclusion of non-standardised measures may, however, affect the reliability/validity of findings and raises the possibility of experimenter bias. Furthermore, a number of measures were delivered through the tablet (Appendix 8.4). Whilst this has the advantage of ease of administration, in studies where a TAU control has been included (Outhwaite et al., 2017-Study 1; Pitchford, 2015; Schacter & Jo, 2016), there is a possibility of bias in favour of the experimental group, who may have greater familiarity with technology at post-test.

2.6.7 Synthesis of key findings

2.6.7.1 Universally-delivered interventions in educational settings

Within educational settings, eight studies have examined the impact of tablet-based mathematics interventions at a universal level, generally reporting positive outcomes (Outhwaite et al., in press, 2017-Study 1; Park et al., 2016; Pitchford, 2015; Schacter & Jo, 2016, 2017, Schacter et al., 2016; van der Ven et al., 2017).

In a preliminary study, Outhwaite et al. (2017-Study 1) examined the impact of the *onebillion* tablet intervention for UK children in the first year of school. They report a significant intervention effect on young children's curriculum knowledge and conceptual understanding of mathematics. Indeed, at post-test, the 4-5 year old children using the intervention were no longer scoring significantly differently on the assessments compared to older peers, aged 5-7 years. Gains

were sustained at 5 month follow-up. The omission of a standardised measure of attainment and an ability-matched control group, however, limit the overall weighting of this evidence to 'medium'.

In a larger scale, follow-up, RCT, Outhwaite et al. (in press) examined the impact of the *onebillion* tablet intervention with 4-5 year old children. This study was the only evaluation to be given a 'high' overall rating within the WoE model. The authors report a significant intervention effect for children using the apps as well as their typical maths lessons (between-group ES: $d = +0.31$) and for children using the apps instead of one regular mathematics activity each day (between-group ES: $d = +0.21$).

Pitchford (2015) used a similar set of tablet apps, also developed by onebillion, for 6-9 year old children in Malawi (WoE rating: 'medium-high'). For children in Standard 2, the second year of school education, the intervention led to significant improvements in the experimental group's curriculum knowledge compared to a TAU control group and a comparison group with access to non-mathematics tablet apps. The intervention also had a significant effect on children's understanding of early mathematical concepts, but only in comparisons between the experimental group and the TAU group. It is possible, therefore, that improvements on this measure were partially due to general access to technology rather than the intervention content. No significant effects were reported for children in Standard 1, the first year of school education, which might be due to the fact that these children used the intervention for 50% less time than the older children.

Three studies conducted in the US evaluated the impact of the intervention Math Shelf (Schacter & Jo, 2016, 2017; Schacter et al., 2016), in a series of evaluations rated 'medium' to 'medium-high' within the WoE model. Across all studies, the intervention had a significant impact on children's overall understanding of number, including quantity discrimination, number recognition and comparison skills (between-group ES: $d = +1.09$, $d = +0.94$, $d = +0.57$, respectively). The magnitude of the ES was lower in Schacter et al.'s (2006) study, which might be due to the shorter duration of the intervention (18

sessions compared with 30/44 sessions). The validity of these studies was, however, weakened by the use of only researcher-developed measures.

Park et al. (2016) determined the impact of a tablet based intervention aimed at developing children's abilities in 'approximate arithmetic', involving adding and subtracting arrays of objects (WoE rating: 'medium'). They reported a significant intervention effect on a standardised measure of general early number skills across the whole sample (between-groups ES: $d = +0.41$). Interestingly, sub-sample analysis showed that the efficacy of the intervention appeared greater for younger children, aged 3-4.9 years.

Finally, van der Ven et al. (2017) evaluated a tablet racing game aimed at developing arithmetic fluency for addition/subtraction facts for 6-7 year old children, see Table 2.7 (WoE rating: 'medium-high'). This intervention appeared to have quite a restricted impact on mathematical attainment, only leading to significant improvements in children's fluency solving subtraction problems involving arrays of dots. There was no significant impact on children's ability to solve dot-addition problems or addition/subtraction problems presented with Arabic symbols. Furthermore, benefits were not sustained at 13 week follow-up. The short duration of the intervention, 5 weeks, may have impacted upon its efficacy.

2.6.7.2 Outcomes for lower attaining children using tablet interventions in educational settings

Only one study has specifically evaluated the impact of a tablet-based intervention on a targeted, low-attaining sample. Outhwaite et al. (2017-Study 4) reported that at-risk 4-5 year old children accessing the *onebillion* intervention made significantly greater gains in curriculum knowledge (within-group ES: $d = +3.3$) and understanding of mathematical concepts (within-group ES: $d = +2.5$) than the control group. ESs were greater than for those reported in Study 1 by the same authors, possibly due to the fact the intervention was implemented for 6 weeks rather than 8 weeks. This study was given an overall 'medium-high' WoE rating, but only a 'low' WoE B rating due to the restricted sample size ($n = 27$) and the lack of equivalence between the initial baseline mathematics skills of lower attaining children receiving the intervention and the higher attaining

control group. There was no significant relationship between children's learning gains and either SES or whether they were learning English as an additional language (EAL), indicating that these factors did not influence outcomes.

In their universal study, discussed above, Outhwaite et al. (in press) reported the impact of the same intervention separately for a sub-sample of low achievers. Within this group they found that learner gains were much greater for children who accessed the maths apps alongside their typical curriculum (within group ES: $d = +4.03$) compared to children in the TAU control group (within group ES: $d = +1.25$). These findings indicate that the *onebillion* intervention may be particularly beneficial for lower-achieving children when used in addition to their typical curriculum. Nevertheless, the authors acknowledge that these findings are only based on a small sample of children and should be treated with caution. It is also important to recognise that differences between groups may also be at least partially attributable to additional exposure to mathematical instruction as opposed to the particular intervention itself.

Researchers evaluating the impact of Math Shelf also explored the efficacy of the intervention for a sub-set of lower achievers in their sample. Schacter et al. (2016) found that the impact of the intervention was greater for children with higher pre-test scores, whilst Schacter and Jo (2016, 2017) found that the intervention was more effective for children with lower pre-test scores. The discrepancy in findings may relate to differences in implementation between the three studies. Lower achievers may have benefited more in the latter evaluations as the intervention was delivered over more sessions and facilitated by a member of teaching staff, who may be more familiar to the children, rather than a researcher.

2.6.7.3 Interventions delivered outside educational settings

Three studies have evaluated the use of tablet-interventions for young children outside of formal education. First, Hieftje et al. (2017) evaluated the efficacy of a mathematics game, Knowledge Battle, during after-school programmes (WoE rating: 'medium'). The authors found a significant effect of the intervention, although it was restricted specifically to children's early number skills, rather

than other aspects of mathematics, including measurement and problem solving.

Kosko and Ferdig (2016) conducted an RCT with 3-5 year old children to evaluate the impact of the *Zorbit* app, which children used at home with varying levels of parent support. They found that the intervention had a significant impact on children's spatial reasoning skills, but not broader aspects of mathematics. Again the short, three week period of the intervention, may have restricted learning gains. Evidence from this study was only rated as 'low' within the WoE model, however, due to the quality and relevance of the research design, see Appendix 8.3.

In another home-based intervention, Berkowitz et al. (2015) found a significant intervention effect from parents reading mathematical stories with their children on an iPad tablet app, over the course of school year (WoE rating: 'medium-low'). It is, however, difficult to determine to what extent effects from these home-based interventions were due to the use of technology per se, rather than increases in the frequency and quality of parent-child interactions around mathematics.

2.6.8 Overall summary of the evidence-base

In general, the review demonstrates that tablet-based interventions can help to raise young children's attainment in mathematics. Positive benefits have been demonstrated across all evaluation studies on at least some outcome measures, although there is considerable variation in ES from small to large (Appendix 8.4). Studies exploring the impact of the intervention for lower achieving children have identified mixed outcomes, although in some cases interventions appeared to more beneficial than for higher achievers (Outhwaite et al., 2018; Schacter & Jo, 2016, 2017). Mixed effects might reflect heterogeneity within these samples and differences in implementation across studies, warranting further exploration. Equally, only one study has found that learning gains were maintained at longitudinal follow-up (Outhwaite et al., 2017-Study 1), highlighting the need for additional research to ascertain whether tablet-based interventions have a sustained benefit.

Several key weaknesses in the evidence-base were identified by this systematic review. First, only two studies have examined the impact of tablet-based interventions for 3-4 year old children (Kosko & Ferdig, 2016; Park et al., 2016). Park et al. (2016) noted, however, that the effects of the intervention were greater for children under 4.9 years in their sample. Given the potential benefits of early intervention, further research is needed to determine whether technology-based programmes might hold the key to ‘closing the gap’ in attainment before children reach school age, providing a rationale for the inclusion of 3-4 year old children in the present study.

Second, the underpinning mechanisms of tablet-based interventions on learning outcomes, and the influence of aspects of implementation, has also not been fully addressed across these studies. The restricted duration of some of the interventions, which lasted only 3-5 weeks, may have limited their efficacy, particularly as there were some indications of increased ES when the same intervention was implemented for longer.

Finally, none of the studies included in this SLR identified the impact of mathematics tablet interventions on young children’s development outside of the domain of mathematics, although the author is aware of some unpublished evidence, as discussed further below, and ongoing investigation in this area. Further evaluation of the wider developmental impact of tablet interventions for young children may be valuable for two reasons. First, it is important to establish whether appropriate use of tablet software can benefit preschool children’s overall development and support other aspects of school readiness. Second, interventions that support other areas of development that continue to influence mathematical learning in the long-term, such as language, may have greater sustained benefits for attainment (see Section 2.7.1).

The following section of the review considers two areas of children’s wider development that may be supported by the *onebillion* tablet intervention, language development and young children’s ‘approaches to learning’, as well as their potential influence on mathematical learning through tablet interventions, providing a rationale for their study in the present research.

2.7 The Potential Cross-Domain Impact of Tablet Interventions

2.7.1 Language development

2.7.1.1 *The importance of early oral language skills*

There is considerable variation in the early language skills of children in the preschool years (Hoff, 2006), with lower language skills often linked with socio-economic disadvantage (Hart & Risley, 1995; Locke, Ginsborg, & Peers, 2002) and, most importantly, children's early communication environment (Roulstone, Law, Rush, Clegg & Peters, 2011). The development of strong oral language skills is thought to provide a gateway to children's learning across all aspects of the curriculum. There is now a body of evidence indicating significant associations between young children's language and later academic achievement, including literacy (e.g. Lepola, Lynch, Kiuru, Laakkonen, & Niemi, 2016; Roth, Speece, & Cooper, 2002) and mathematics (e.g. LeFevre et al., 2010; Praet et al., 2013; Purpura & Ganley, 2014, Vukovic & Lesaux, 2013). The importance of early communication and language skills is reflected in the UK EYFS curriculum, where they are identified as one of three prime areas of learning (DfE, 2017b).

A number of specific components of language may be particularly important for mathematical development. First, an extensive range of mathematics-specific vocabulary may be required. Counting and number recognition skills, for instance, depend on knowledge of number names, whilst understanding vocabulary such as "more than", "under", "longer" and "next to" help children to develop mathematical concepts of comparison and measurement (Powell & Driver, 2015; Toll & Van Luit, 2014). Studies show that the amount of mathematics 'talk' that children hear from their preschool teachers and parents is associated with their mathematical ability (Gunderson & Levine, 2011; Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006; Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010; Susperreguy & Davis-Kean, 2016). Moreover, children's understanding of mathematical language is a predictor of future numeracy skills (Purpura & Reid, 2016; Toll & Van Luit, 2014).

Other aspects of language, however, may also play an important role in mathematical learning. A recent study, conducted by Chow and Ekholm (in press), for instance, has shown that 6-8 year old children's receptive syntactic skills (understanding of word order and combinations) are more strongly associated with concurrent mathematical attainment than general vocabulary knowledge. Syntactic ability may help children to understand grammatically complex instructions in the classroom and solve worded problems. Whilst further longitudinal research is required to determine how different components of language ability might mediate growth in mathematical attainment over time, it is, nevertheless, clear that early receptive language skills play an important role in mathematics and broader learning.

2.7.1.2 *Educational technology and language skills*

Whilst there has been no published research as yet exploring the impact of tablet-based mathematics interventions on children's language development, Sarama et al. (2012) identified that the early mathematics intervention, *Building Blocks*, which has a computerised component, benefits 5-6 year old children's oral language skills.

There are a number of different mechanisms by which the *onebillion* intervention might similarly promote early language development. First, the app software places an emphasis on modelling a variety of mathematics-specific vocabulary, including prepositional language ('above', 'next to') and language related to quantity ('more than', 'less than'). Direct teaching and modelling of vocabulary has been shown to improve early language development (Bickford-Smith, Wijayatilake, & Woods, 2005). Second, as children follow directions from a virtual teacher, they gain increased practice in following instructions, varying in terms of length and grammatical complexity. It is therefore hypothesised that the *onebillion* software may hold benefits for children's understanding of vocabulary and syntax, which in turn may further benefit mathematical learning.

It is, however, important to note that children's initial language skills could also potentially moderate the gains that children make whilst using the app software. For example, given the language demands of listening to instructions, children with poorer language skills may not be able to fully access app content.

Alternatively, children with weaker language skills may demonstrate greater learning gains if use of the intervention enables them to overcome a previous barrier to their mathematical learning. Preliminary research with 4-5 year old children, however, indicates that receptive vocabulary and EAL status do not affect learning gains from using the *onebillion* intervention (Outhwaite et al., 2017). The present research seeks to address these relationships further with younger learners, using a broader assessment of receptive language skills.

2.7.2 Approaches to learning

Reviewers have also called for greater emphasis on promoting children's "persistence, self-control, and curiosity" (Gersten, 2016, p. 687) during early years mathematics interventions. It is commonly recognised that there are considerable differences in the way that young children approach learning activities which they encounter (Chen & McNamee, 2011). For instance, children may differ in their enthusiasm and persistence when solving puzzles, or the interest that they show in new educational toys. Approaches to learning (AtL) are thought to be important due to the influence they may have on early academic achievement (Kagan, Moore & Bredekamp, 1995); it is arguable that a child who is goal-orientated and persists for longer at an educational activity is likely to make greater learning gains than a child who is hesitant, distractible or less organised (Chen & McNamee, 2011).

2.7.2.1 Defining 'approaches to learning'

AtL, a term introduced to capture the notion of developmental receptiveness to learning environments, can be defined as "observable behaviours that describe ways children engage in classroom interactions and learning activities" (Chen & McNamee, 2011, p.78). Under the umbrella construct of AtL, researchers have explored a number of different learning behaviours which are thought to be important in the early years, including curiosity, initiative, task persistence, attentiveness, engagement, organisation and flexibility in problem-solving (Barbu, Yaden, Levine-Donnerstein, & Marx, 2015; Chen & McNamee, 2011; DiPerna, Lei, & Reid, 2007; Li-Grining, Votruba-Drzal, Maldonado-Carreno, & Haas, 2010). In part, AtL can be considered the behavioural manifestation of children's executive function skills (higher order cognitive control processes),

such as attention, working memory and inhibitory control; however they also reflect the development of social-emotional competencies, such as independence and responsibility (McClelland, Acock, & Morrison, 2006). AtL can therefore be viewed as being at the centre of the interaction between young children's social, emotional and cognitive development and the knowledge that they are able to acquire from the world around them.

2.7.2.2 The importance of positive learning approaches for academic achievement

Longitudinal research suggests that individual differences in young children's AtL are predictive of both concurrent and future academic achievement in reading and mathematics (Ansari & Gershoff, 2015; DiPerna et al., 2007; Li-Grining et al., 2010; McClelland et al., 2006; McClelland, Morrison, & Holmes, 2000; Razza, Martin, & Brooks-Gunn, 2015). Studying a large nationally representative sample of children in the United States, Li-Grining et al. (2010) found that 5-6 year old children with more positive AtL, measured by teacher ratings of classroom behaviour, experienced a faster rate of growth in their reading and mathematics achievement through to the age of 10-11 years, after controlling for initial skill levels. They also found that early AtL appeared to be more beneficial for the attainment of children who began school with lower academic skills. AtL may therefore be a protective factor for young children with low early attainment, enabling them to capitalise more on the learning opportunities which they experience at home and school (Li-Grining et al., 2010).

2.7.2.3 Educational technology and approaches to learning

It is likely that more positive AtL can be promoted in young children through the learning activities that they access. The particular characteristics of an activity, including its goals, materials and the cognitive demands placed on the child, are likely to affect not only *what* a child learns but the *way* that they learn (Chen & McNamee, 2011). Use of the *onebillion* tablet intervention requires children to persist with problems that they encounter, flexibly change their problem-solving approach, attend carefully to a virtual teacher and work independently, which might give them an opportunity to develop each of these aspects of AtL. Moreover, motivational and engaging features of the software, including visual

images, characters and positive rewards (e.g. certificates), and the goal-directed nature of the intervention may increase engagement and on-task behaviour. This assertion is supported by evidence that use of the *onebillion* intervention, rather than other game-based tablet apps, significantly benefits 6-9 year old children's selective attention skills, as measured by a visual search task (Pitchford & Outhwaite, in prep). The present research aims to build upon these findings, exploring the possible impact on observed classroom learning behaviours in preschool children.

The potential benefits of tablet technology for AtL are also noted by Course and Chen (2010) who conducted structured observations of 3-6 year old children whilst they were using a tablet drawing app. During the activity, children became increasingly independent and persisted even when they encountered technical problems. Moreover, in a qualitative study, Clarke and Abbott (2016) note perceived benefits in the development of children's confidence from using tablet apps. The relationship between young children's AtL and tablet-based interventions was not, however, quantitatively evaluated in these studies.

Teachers are also thought to have an important role in fostering more positive AtL (Ansari & Gershoff, 2015; Williford, Maier, Downer, Pianta, & Howes, 2013). Whilst facilitating children's use of the apps, teachers may be able to promote their AtL by providing specific instruction, coaching and modelling during the activity (e.g. "*Listen carefully*") and by helping to children to reflect on their learning, particularly when they encounter problems (e.g. "*Is there another way you can try that?*") (Chen & McNamee, 2011).

The present research therefore aimed to determine whether use of a tablet-based mathematics intervention, facilitated by teaching staff, might promote more positive AtL. It also considers whether children's initial AtL might affect the extent to which children benefit from using the intervention. Hypotheses were bidirectional given the exploratory nature of the research; it is possible, for example that some children might not be able to benefit as fully from the intervention until they have learnt "*how to learn in a structured environment*" (Ansari & Gershoff, 2015, p. 700). Indeed, Pitchford et al. (2018) found that in a small sample of children with special educational needs and disabilities (SEND),

task engagement predicted the progress rate of children during the *onebillion* intervention. Alternatively, children with lower AtL may demonstrate greater learning gains if using the intervention helps them to develop more positive learning behaviours and removes barriers to learning.

2.8 The Aims and Original Contribution of the Proposed Research

2.8.1 Raising mathematical attainment in the preschool years

In light of the theory and literature considered in this review, the primary aim of the current study was to extend the evidence-base by evaluating whether the *onebillion* maths tablet intervention is effective for younger children, aged 3-4 years and attending Foundation Stage 1 nursery classes. The study also aimed to address the methodological limitations of previous research, identified in the SLR, by implementing the intervention over a 9 week period, incorporating a 5 month longitudinal follow-up and ensuring that a standardised measure of mathematics is included in the design.

2.8.2 The cross-domain impact of the *onebillion* intervention

A further unique contribution of this research was to explore whether the *onebillion* intervention has additional benefits for 3-4 year old children's broader development and school readiness. The current research aimed to determine the potential impact of using the apps on young children's receptive language skills and teacher-rated AtL.

2.8.3 Factors affecting intervention outcomes

A final aim of the research was to illuminate the mechanisms underpinning the efficacy of the *onebillion* intervention with preschool children. Through observations of children's use of the intervention and semi-structured interviews with facilitators, the present research aimed to elucidate how characteristics of the children, software features, and aspects of implementation (e.g. facilitator support and wider class/school factors) may impact upon learning outcomes. A more exploratory view was also taken of possible outcomes of the intervention by considering facilitators' perceptions of intervention effects.

Additional quantitative analyses was undertaken to consider whether learning gains are affected by initial ability, SES, receptive language skills and AtL, exploring whether any of these factors might affect potential individual differences in Rtl. Greater understanding in this area may hold implications for the development of app software and effective implementation of tablet technology in early education.

2.9 Research Questions

The present study, therefore, aims to answer the following research questions:

- 1.) What is the impact of the *onebillion* tablet intervention on the mathematical attainment of preschool children aged 3-4 years?
- 2.) Are mathematical attainment gains sustained 5 months after the end of the intervention?
- 3.) What is the impact of the *onebillion* tablet intervention on the receptive language skills of preschool children aged 3-4 years?
- 4.) What is the impact of the *onebillion* tablet intervention on the AtL of preschool children aged 3-4 years?
- 5.) What are facilitators' perceptions of the outcomes of the *onebillion* intervention for preschool children aged 3-4 years?
- 6.) What factors may affect the outcomes of the *onebillion* intervention for preschool children aged 3-4 years?

3 Methodology

This chapter begins with a discussion of different research paradigms used within psychological and educational research in order to contextualise the methodology of the present study. The research design, sampling procedures, data collection methods and implementation procedures are then presented.

3.1 Paradigms within Applied Educational Research

Research paradigms, or worldviews, reflect the accepted shared beliefs, values and practices of different groups of researchers (Creswell, 2014; Morgan, 2014). They differ according to the following philosophical assumptions:

- **Ontology:** beliefs about the nature of reality
- **Epistemology:** beliefs about the nature of knowledge and the relationship between the researcher and the knowledge that they hope to obtain
- **Methodology:** the set of approaches thought to be suitable in obtaining desired knowledge and understanding

(Mertens, 2015; Morgan, 2007)

It is important for the researcher to have an awareness of these assumptions, given that they will influence the decisions made throughout the research process (Fielzer, 2010; Mertens, 2015; Ponterotto, 2005).

3.1.1 Alternative paradigms

Post-positivism and constructivism, two dominant paradigms in psychological and educational research, have traditionally been viewed as fundamentally opposing positions (Feilzer, 2010). Post-positivists assume the existence of a single external reality, which can be understood through direct experience and observation of the world (Robson, 2011). Post-positivist researchers, however, maintain that this reality can only be known imperfectly and probabilistically, as there will always be limitations to the reliability and validity of evidence obtained (Ponterotto, 2005). In order to increase objectivity, post-positivist research

typically involves quantitative, controlled designs to test the likelihood of particular hypotheses and theories (Robson, 2011).

In direct contrast, constructivism holds that reality is socially constructed by human beings and that there is no single objective truth (Ponterotto, 2005). A researcher's role is to seek greater understanding of the multiple meanings and viewpoints that might exist about a particular phenomenon (Mertens, 2015). Instead of hypothesis testing, constructivist research typically involves qualitative designs in order to: illuminate the perspectives and lived experiences of different individuals, identify patterns of meaning and, in turn, generate new theories through inductive reasoning (Creswell, 2014).

More recently, pragmatism has emerged as an alternative paradigm, potentially reconciling differences between the ontological and epistemological assumptions of earlier paradigms. Pragmatists maintain that a single external reality exists, but also accept that all individuals will have their own individual interpretation of this reality (Mertens, 2015). Knowledge is therefore both constructed and based upon the external world (Johnson & Onwuegbuzie, 2004). The value of research is determined by its effectiveness in answering a particular research question and solving practical real-world problems, rather than in seeking one particular type of truth (Fielzer, 2010; Johnson & Onwuegbuzie, 2004). Pragmatism often underpins mixed methods research, combining the use of quantitative and qualitative approaches depending upon 'what works' in achieving the researcher's purpose (Biesta, 2010).

3.1.2 The philosophical assumptions underpinning the current study

A pragmatic stance was adopted by the researcher, reflected in the use of a mixed methods design. Research questions evaluating the causal impact of the intervention were addressed through quantitative approaches and an experimental design. Additional qualitative forms of data were gathered, however, to gain further insight into aspects of implementation and factors affecting outcomes, exploring facilitators' viewpoints and children's experiences. In combining methodological approaches to answer different research questions, this study draws upon the philosophical assumptions of both post-positivism and constructivism.

3.2 Stakeholders

A number of key stakeholders were considered in the planning of this research:

- Participating children, their parents and nursery-school staff
- The researcher
- The UoN

A team of researchers at the university have conducted previous evaluations of the onebillion intervention. Collaboration within this research team has therefore been essential in supporting the development and implementation of this research.

- The EPS

As noted above, the researcher holds a bursary training placement within an EPS. It was therefore important that the research met service priorities and was beneficial for the children and young people within the Local Authority where the EPS is located.

- The educational not-for-profit organisation, onebillion

It is important to acknowledge that it was necessary to discuss the present research with the developers of the intervention, onebillion, who provided the software free of charge to all schools participating in the research.

- The wider educational and psychological research community

Discussions with all key stakeholders were conducted during the planning and setup of this research to ensure that the research met their goals, expectations and was feasible to implement.

3.3 Mixed Methods Research Design

3.3.1 Mixed methods and applied research

In line with the pragmatic paradigm, mixed methodologists seek to combine quantitative and qualitative data collection depending upon what works best in answering the researcher's question(s). Mixed methods research has been defined as:

"...the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g. use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration"

(Johnson, Onwuegbuzie, & Turner, 2007, p. 123)

Instead of assuming a dichotomy between quantitative and qualitative approaches, Teddlie and Tashakkori (2009) proposed that different research traditions lie on a continuum, as displayed in Figure 3.1. The left circle represents a typical qualitative approach where the researcher is seeking to address exploratory research questions through the use of inductive reasoning. In contrast, the right circle represents the quantitative tradition, where the researcher is addressing confirmatory research questions, applying deductive logic and inferential statistics to test particular hypotheses. Mixed methodologists are seated in the centre, combining approaches and potentially varying their position within the exploratory-confirmatory continuum at different stages in the research process (Teddlie & Tashakkori, 2009).

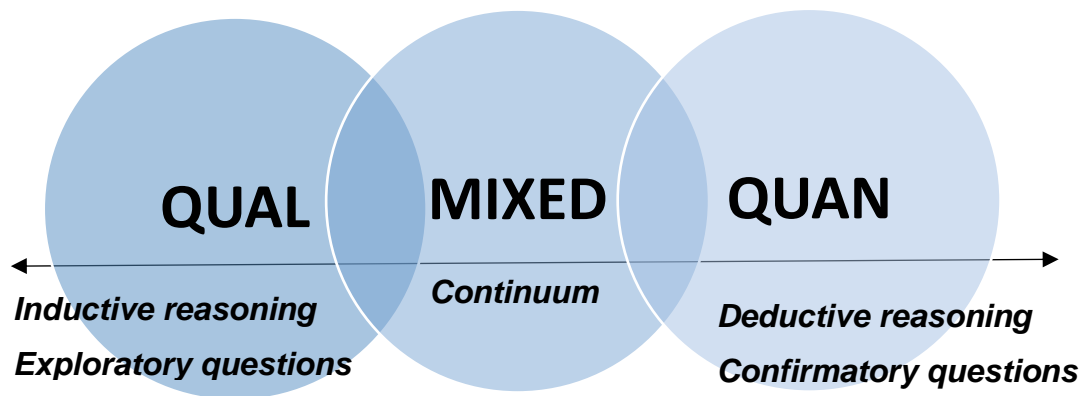


Figure 3.1: The Qualitative-Mixed Methods-Quantitative (QUAL-MM-QUAN) continuum (adapted from Teddlie and Tashakkori, 2009, p.28)

A variety of different mixed methods designs have been used within applied research, differing according to (a) the purpose for combining methodologies, (b) the dominance of any methodological approach (quantitative or qualitative) and (c) the point at which different types of data are collected and analysed (Bryman, 2006). Creswell (2014) has identified a number of common mixed methods research designs:

- A convergent parallel mixed methods design:
The researcher collects both quantitative and qualitative data at the same stage in the research study. Both sets of data are analysed and integrated in order to provide an overall answer to the research problem, highlighting discrepancies where necessary.
- Explanatory sequential mixed methods design:
This type of research begins with a quantitative research phase, followed by a qualitative phase which aims to explain the findings in more detail. For example, a researcher might conduct a survey, analyse the data and then select participants for follow-up qualitative interviews.
- Exploratory sequential mixed methods design:
The researcher begins with a qualitative research phase to explore the views of participants. Analysis from this phase is then used to inform the

nature of a subsequent quantitative research phase. For example, focus groups might be used to determine variables of interest in an experimental design.

- Embedded mixed methods design:
In this design, one or more sets of data (quantitative and/or qualitative) is nested within a larger design, such as an experiment. Supplementary data can be collected either before, during or after the main phase of data collection.

3.3.2 Research design of the current study

The present research used an embedded mixed methods design to address the research questions stated in Section 2.9. The dominant approach to data collection was quantitative, using a quasi-experimental design (with longitudinal follow-up) to answer causal questions about the efficacy of the *onebillion* intervention.

In order to explain the experimental findings in more depth, supplementary data were gathered both during and following the intervention. The purpose of gathering this additional data was twofold. First, the researcher aimed to take a more exploratory view of the possible outcomes of the intervention by illuminating facilitators' perceptions at the end of the intervention and triangulating these with the experimental findings. Second, the researcher hoped to provide greater insight into factors affecting the outcomes of the intervention, including pupil characteristics, particular features of the intervention and broader class/school level factors.

The following additional exploratory forms of data were therefore collected:

- Semi-structured qualitative interviews with facilitators at post-test to gather their views
- Narrative qualitative observations of the intervention sessions conducted by the researcher

- Quantitative structured observations of children's attention on-task during intervention sessions to explore children's engagement over the course of the intervention period
- Quantitative analysis of associations between particular characteristics of participants (including SES, initial maths attainment, receptive language and AtL) and gains made through the intervention.

The study was conducted across three key time points, as illustrated in Figure 3.2 and discussed further across Sections 3.4 and 3.5.

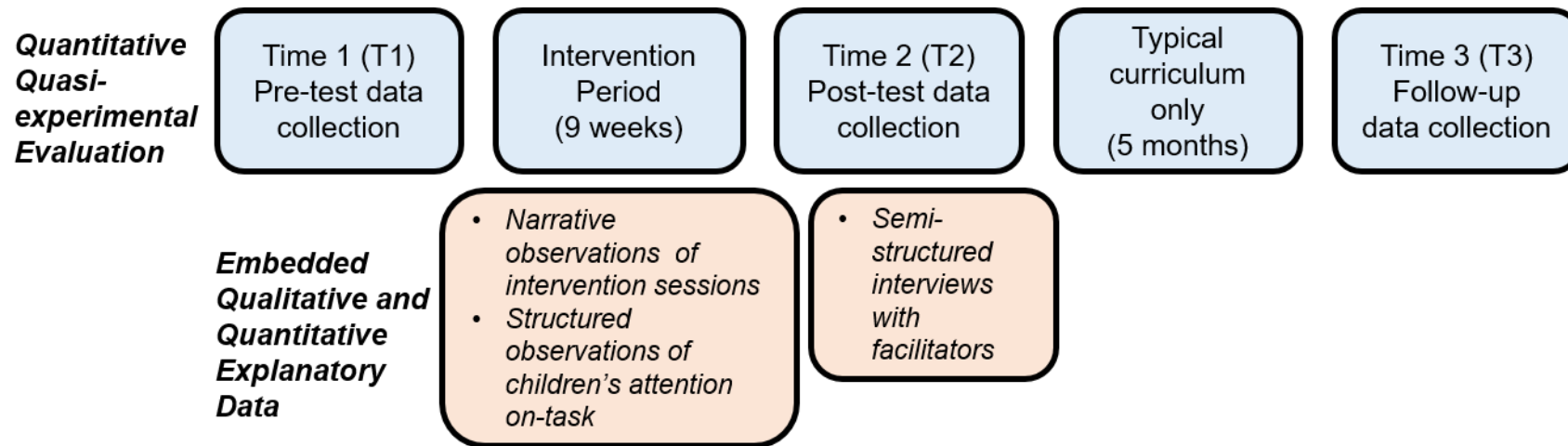


Figure 3.2: An illustration of the embedded mixed methods design

3.4 Quantitative Evaluation of the Efficacy of the Intervention

3.4.1 Research Questions 1-4 and associated hypotheses

The purpose of the quantitative evaluation was to test causal hypotheses (Table 3.1) regarding the efficacy of the *onebillion* tablet intervention in developing young children's mathematical attainment, receptive language skills and positive AtL.

Table 3.1: Research questions and hypotheses for the quasi-experimental design

Research Question	Experimental Hypothesis	Null Hypothesis
1. What is the impact of the <i>onebillion</i> tablet intervention on the mathematical attainment of preschool children aged 3-4 years?	Using the <i>onebillion</i> tablet intervention has a statistically significant effect on 3-4 year old children's mathematical attainment at post-test.	At post-test, any differences between groups in mathematical attainment will be due to chance.
2. Are gains in mathematical attainment sustained 5 months after the end of the intervention?	Using the <i>onebillion</i> tablet intervention has a statistically significant effect on 3-4 year old children's mathematical attainment 5 months after the end of the intervention.	At 5 month follow-up, any differences between groups in mathematical attainment will be due to chance.
3. What is the impact of the <i>onebillion</i> tablet intervention on the receptive language skills of preschool children aged 3-4 years?	Using the <i>onebillion</i> tablet intervention has a statistically significant effect on 3-4 year old children's receptive language skills at post-test.	At post-test, any differences between groups in receptive language will be due to chance.
4. What is the impact of the <i>onebillion</i> tablet intervention on the AtL of preschool children aged 3-4 years?	Using the <i>onebillion</i> tablet intervention has a statistically significant effect on 3-4 year old children's AtL at post-test.	At post-test, any differences between groups in AtL will be due to chance.

3.4.2 Evaluation design

Alternative research designs that were considered to evaluate the impact of the *onebillion* tablet intervention are discussed here, presenting a rationale for the choice of evaluation design used in the present research.

3.4.2.1 *Randomised control trials*

RCTs, often regarded as the 'design of choice' for evaluating educational interventions (Slavin, 2002), involve inclusion of a control condition to establish whether any changes in the experimental condition are due to children accessing the intervention, rather than history, maturation, testing effects or other extraneous factors (Mertens, 2015). Moreover, randomisation of participants to condition reduces the possible threat of selection bias and increases the likelihood that groups will be equivalent at baseline (Robson, 2011; Slavin, 2002). These features allow the researcher to make stronger causal inferences about the impact of the intervention, known as internal validity. Consequently, findings from RCTs are typically ranked highly within the 'hierarchy of evidence' (Petticrew & Roberts, 2006).

3.4.2.2 *Matched designs*

In real world research, it is often difficult to obtain a large enough sample size to ensure that randomisation will lead to baseline equivalence across groups (Robson, 2011). In evaluations of educational interventions, it is particularly important that there are no significant differences between the experimental and control groups in the spread of children's ability at pre-test on key dependent variable (DVs) (Torgerson & Torgerson, 2003). Children with different baseline levels of attainment have the potential to make differing amounts of progress over the course of the experimental period, due in part to the measurement properties of any assessment as well as the natural trajectory of development for those learning skills. If the spread of ability across both conditions is not equivalent at pre-test, then differences in 'potential' learning gains will be confounded with any improvements caused by the intervention.

In an RCT, remaining between-group differences in pre-test ability can be controlled to an extent during statistical analysis (EEF, 2015). It is arguably better to also minimise any potential group differences through the choice of

research design (Rubin, 2008). One possibility is to use ability-matching, where a similar number of high, low and medium attaining children are allocated to control and experimental groups (Creswell, 2014). It is important to note, however, that the accuracy of matching will necessarily depend upon the reliability and validity of measurement. The researcher also has to make a choice about which participant characteristics are most appropriate for matching and acknowledge that some group differences may remain.

3.4.2.3 Choice of design for the current study

In this research, an ability-matched pre-test post-test quasi-experimental design (with 5 month follow-up) was used to evaluate the impact of the *onebillion* tablet intervention used by 3-4 year old children attending nursery school. The independent variable (IV) was the condition which the children were allocated to and the DVs were the outcomes on the children's mathematics attainment, receptive language and AtL.

A pre-test post-test controlled design was selected in order to evaluate the efficacy of the *onebillion* intervention and to ensure strong internal validity. The sample size of this study was restricted, however, due to practical constraints in sampling. Therefore, in order to achieve equivalence across experimental and control conditions at baseline, ability matching was used to allocate participants to condition rather than randomisation. Pre-test mathematics attainment was selected as the matching variable, given that children's gains in mathematics were the primary focus. Children in each class were ranked according to their pre-test mathematics scores and then alternately allocated to control and experimental groups (see Section 3.4.6). In order to determine whether learning gains in mathematics were sustained, a longitudinal follow-up was conducted 5 months after post-test.

3.4.3 Sampling

3.4.3.1 Nursery selection

The researcher sent an 'Expression of Interest' letter (Appendix 8.5) to all schools with nursery classes in a town in the West Midlands. Only school-based nurseries were contacted in order to increase the likelihood that the setting would already have access to Apple iPad tablet devices and also to facilitate

longitudinal follow up 5 months later, as many children were likely to remain at the same setting when they started school. In order to participate, nurseries had to meet the following selection criteria:

- Typical attendance at the nursery was on a daily basis (five half-day sessions per week)
- At least five school-owned iPad tablet devices were available for 15 minutes a day during the 9 week intervention period
- Nursery staff were available on a daily basis to facilitate the intervention

Four schools expressed interest in participating in the research and meetings were held with nursery staff to discuss the possibility of involvement and to ensure that they met selection criteria. Following these meetings, two schools were keen to participate. Head teachers of these schools were provided with an information sheet, outlining the responsibilities of the school and the researcher, and gave written consent for school involvement in the research (Appendix 8.6). Demographic details of these schools are shown in Table 3.2.

Table 3.2: Demographic details of participating schools⁶

School Type and Ofsted Rating⁷	Total Number of Pupils (to the nearest 50)	% with a Statement of SEN or EHCP⁸	% EAL⁹	% Eligible for FSM¹⁰	Nursery Sessions Times (Number of pupils)
School A Academy <i>Good</i>	500	1.2	19.3	17.1	Mornings (<i>n</i> = 19) Afternoons (<i>n</i> = 10)
School B Academy <i>Inadequate</i>	200	0.5	31.6	40.2	Mornings (<i>n</i> = 22)

3.4.3.2 Pupil selection

All 3-4 year old children attending Foundation Stage 1 at the two nurseries were invited to participate in the research. Parents of all pupils were invited to attend information sessions at each school where the researcher outlined the purpose and nature of the project and obtained written consent (Appendix 8.7).

Information sheets and consent forms were also distributed to parents who could not attend the meeting. Assent was also obtained from the children to participate in the individual assessments (Appendix 8.8). Figure 3.3 shows a flowchart demonstrating the process of pupil selection. In total, 47 children participated in the research, although it was not possible to obtain a full data set for all children, as discussed in Section 4.2.

⁶ Data taken from Department for Education (2017) School and College Performance Tables (<https://www.compare-school-performance.service.gov.uk>, accessed 21.01.18)

⁷ Both schools had recently converted to academies and therefore Ofsted judgements reported here are those for the predecessor schools.

⁸ Percentage of Pupils with a Statement of Special Educational Needs or Education, Health and Care Plan, where the national average was 3%.

⁹ Percentage of pupils whose first language is not English, where the national average was 20.5%.

¹⁰ Percentage of pupils eligible for free school meals at any time during the past 6 years (proxy measure of SES), where the national average was 24.7%.

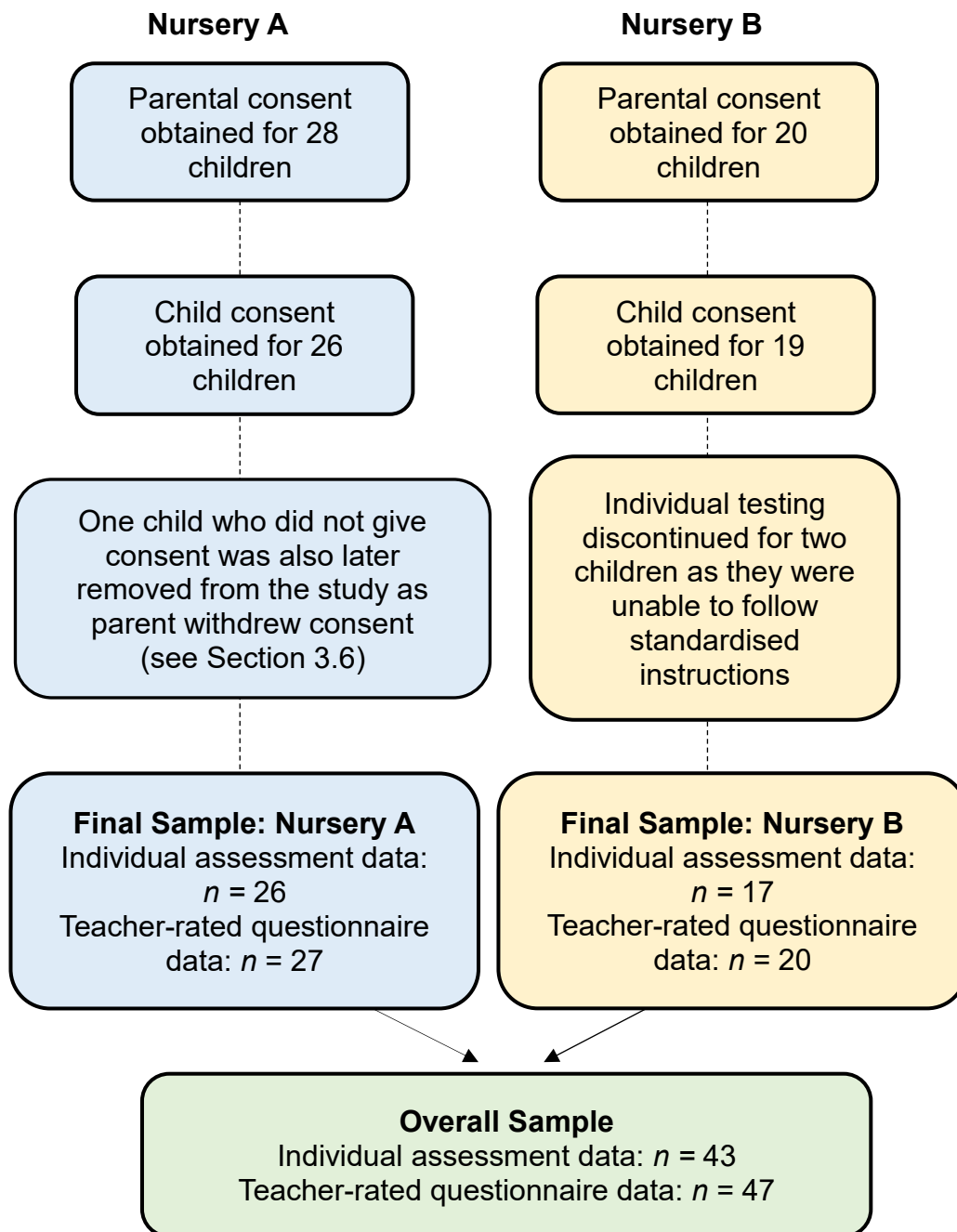


Figure 3.3: Pupil sampling flowchart

Demographic data for participating children are outlined in Table 3.3 below. Across the sample there was a relatively high proportion of children with EAL and/or low SES, particularly within Nursery B. Teachers reported that children at Nursery A did not use tablets at the setting prior to the study. Children at Nursery B, however, had regular access to iPad tablets (2-3 times per week), including other maths apps, in their classroom environment to use during periods of self-directed play.

Table 3.3: Demographic details of participating sample

Setting	Mean Age in Months (Standard deviation)	Gender (% male)	SES (% IDACI Decile Rank ¹¹)				EAL Status (%)
			≤ 3	4 - 6	≥ 7	ND ¹²	
Nursery A (<i>n</i> = 27)	48.33 (3.31)	59.3	63.0	18.5	11.1	3.7	11.1
Nursery B (<i>n</i> = 20)	48.70 (4.03)	55.0	85.0	10.0	0.0	5.0	40.0
Total 47	48.49 (3.59)	57.4	68.1	14.9	6.4	4.3	23.4

¹¹ IDACI = Income Deprivation Affecting Children Index. IDACI scores are a measure of children's socio-economic status, calculated from 2015 UK government survey data about the level of deprivation within a given locality (Department for Communities & Local Government, 2015).

IDACI scores are given a decile rank from 1-10: a locality with a decile score of 1 falls among the top 10% most deprived areas in England.

Children's postcodes were used to calculate IDACI scores using the following website <http://imd-by-postcode.opendatacommunities.org/> (accessed 25.01.18).

¹² ND = postcode not disclosed by parents

3.4.4 Conditions

Pupils were allocated to either an experimental condition or to a wait-list control condition, as detailed below.

3.4.4.1 *Experimental condition*

Children allocated to the experimental condition used the *onebillion* tablet intervention for 15 minutes each day for 9 weeks (45 sessions) during their typical nursery sessions. As noted previously, the intervention consists of two separate apps, *Maths Age 3-5* and *Maths Age 4-6*, treated for research purposes as a single continuous intervention. All children accessing the intervention started with the activities at the beginning of the *Maths Age 3-5* app, progressing through topics in the same order but at their own pace. The intervention was delivered in small groups within the child's normal classroom, facilitated by nursery staff (see Section 3.4.6). Children accessed the apps individually on touchscreen Apple iPad tablet devices and used headphones to minimise noise disturbance. Children received the intervention *in addition* to their typical instruction in mathematics, as discussed below.

3.4.4.2 *Control condition*

Children in the TAU control condition did not access the intervention during the experimental period. These children typically engaged in self-directed play whilst the intervention was used by the experimental group. However, they continued to receive their typical mathematics instruction, see below, and had regular access to iPad tablets for other purposes during their nursery sessions.

3.4.4.3 *Typical Instruction in Mathematics*

Drawing upon observations (conducted prior to the start of the project) and teacher reports, Table 3.4 shows the typical type and frequency of mathematics instruction that children in both groups received across Nursery A and Nursery B. It is acknowledged, however, that the nature of the wider teaching of mathematics in the nursery classes was not fully audited and monitored over the course of the intervention period. Nevertheless, during intervention fidelity checks (see Appendix 8.11) teachers reported that both groups continued to receive their typical instruction in mathematics.

Table 3.4: Typical mathematics instruction across settings

Nursery	Typical Mathematics Instruction
Nursery A	<ul style="list-style-type: none"> • Whole class teaching of mathematics (lasting approximately 15 minutes, typically once per week) followed by linked adult-led activities to extend learning, e.g. use of <i>Numicon</i> materials (see Oxford University Press, 2018) • Maths activities setup in the class environment that children could self-select during play-based learning sessions (e.g. large dice available for children to roll) • Daily quick maths class activities (e.g. counting) • Additional 1:1 support for children if needed for counting and shape/number recognition
Nursery B	<ul style="list-style-type: none"> • Whole class teaching of mathematics (lasting approximately 10 minutes, typically twice per week) (e.g. counting songs) • One to one sessions for each child to extend their learning based on identified next steps, e.g. using small toys for learning counting principles and understanding one more/one less (typically lasting 2-5 minutes, twice per week) • Adult-led focus group activities (lasting approximately 10 minutes, typically once per week) • Maths activities setup within the class environment that children could self-select during play-based learning sessions

3.4.5 Measures

Measures used to assess each DV within the quantitative evaluation phase of the research design are outlined below, together with further detail about the administration of each measure.

3.4.5.1 Mathematics attainment measures

3.4.5.1.1 Standardised measure

In order to increase the reliability and validity of the design, a standardised measure was chosen to assess mathematics attainment over time and to ensure that groups were appropriately matched for pre-test ability. Table 3.5 details the mathematics measures identified by the researcher that have been

standardised for use with 3-4 year old children and the factors considered in test selection.

The Early Number Concepts (ENC) sub-test was selected from the British Ability Scales (BAS)-III (Elliot & Smith, 2011) as it was the only measure standardised in the UK (see Table 3.5). The ENC sub-test has high internal reliability (Cronbach's $\alpha = .94/.95$) and test-retest reliability ($r = .81$). The scale comprises 30 items, measuring a number of early maths skills and concepts, including: counting skills, size matching, number recognition and understanding of vocabulary (e.g. *more* or *less*).

The scale takes approximately 10 minutes to administer individually to each child, reducing demands on the sustained attention of young children. Whilst the format of the questions varies, children are asked to respond verbally and/or by pointing to multiple-choice picture answers in the test booklet. Children start and finish the assessment at different points depending upon their age, although earlier and later items may be administered if children pass fewer than 3 items or make fewer than 3 mistakes, respectively. Raw scores are therefore adjusted depending on the start/end points to give an ability score (used in subsequent analyses).

Table 3.5: Standardised mathematics measures for 3-4 year old children

Name of Measure	Source	Age Range	Approximate Administration Time	Parallel test forms	Type of mathematics assessed	Standardised in the UK
Early Number Concepts sub-test from the BAS - III	Elliot & Smith (2011)	3:00-7:11 years	10 minutes	No	Early number concepts, including: counting, matching according to size, matching according to number, addition, number recognition, understanding of most, one more and one less	Yes
Test of Early Mathematical Ability-III	Ginsburg & Baroody (2003)	3:00-8:11 years	40 minutes	Yes	Measures concepts and skills in the area of number, magnitude comparison, numeral literacy, basic facts and calculation skills	No
Woodcock-Johnson IV Tests of Early Cognitive and Academic Development – Number Sense Scale	Schrank, McGrew & Mather (2015)	2:06-7:11 years	Unknown	No	Measures number sense, including: number recognition, spatial/size orientation, counting, number line estimation, number sequencing, magnitude representation and inductive reasoning	No

3.4.5.1.2 *Researcher-designed measure*

The ENC sub-test is not, however, tightly linked to the EYFS curriculum for mathematics in the UK and therefore may not be sufficiently sensitive to intervention learning gains. Consequently, a researcher-designed measure of mathematical curriculum knowledge (CK) was also selected, developed by Outhwaite et al. (2017) (Appendix 8.9). The test includes 50 novel test-items based on the questions and concepts taught in the maths app, including number recognition, counting, shape recognition and basic addition/subtraction. It has two parallel forms to reduce the possibility of practice effects. The questions are designed to reduce demands on young children's language and memory skills. The CK assessment was administered orally by the researcher on an individual basis. Children responded either orally or using paper/pen, depending on their fine motor skills.

Use of the CK measure supports comparison with past research as this measure has been used in previous evaluations of the intervention with 4-5 year old children (Outhwaite et al., 2017). However, as the children were younger in the present research, a discontinue rule was introduced if children made more than 5 consecutive errors. Raw scores were used in the analyses.

The scale has not been standardised and so the reliability and validity of the measure was assessed in the course of the research. The internal consistency of the scale was high at T1 (Kuder-Richardson 20 values were .91 and .95 for Forms A and B respectively). Test-retest reliability between T1 and T2 (across alternate test forms) was also high ($r = .83$)¹³. There was also a strong correlation between scores on the CK assessment and ENC assessment at T1, indicating good criterion-related reliability ($r = .75$).

3.4.5.2 *Receptive language*

A standardised measure was also selected to assess children's receptive language. Measures that were considered for inclusion in the study and which

¹³ Pearson's Product Moment Correlation Coefficient was calculated for the control group only to determine test-retest reliability, given that use of the intervention was hypothesised to affect these scores.

have been standardised for use with 3- 4 year old children are included in Table 3.6 below.

The Clinical Evaluation of Language Fundamentals - Preschool 2 UK (CELF-P2) (Wigg, Secord & Semel, 2006) contains three sub-tests that collectively give an index of receptive language skills: Sentence Structure, Basic Concepts, and Concepts and Following Directions. This measure was selected as it directly assesses children's understanding of many of the specific semantic concepts taught by the app (e.g. number, size, position). The Concepts and Following Directions scale also measures children's ability to follow instructions of increasing length and syntactic complexity. This assessment has high test-retest reliability ($r = .92-.95$) and internal consistency (Cronbach's alpha = $.91-.94$). The subscales are administered orally on an individual basis and children respond non-verbally by pointing at multiple choice picture answers. A total raw score was calculated across scales in this measure and used for analyses.

Table 3.6: Standardised receptive language measures for 3-4 year old children

Name of Measure	Source	Age Range	Approximate Administration Time	Language skills assessed
Verbal Comprehension sub-scale – BAS - III	Elliot & Smith (2011)	3:00-7:11 years	10 mins	Understanding of oral language instructions (including one and two step instructions) Understanding of basic language concepts (including body parts, objects and prepositions)
CELF-P2 UK	Wiig et al. (2006)	3:00-6:00 years	15 mins	<ul style="list-style-type: none"> • Sentence structure sub-test: assesses comprehension of sentence formation rules and understanding of sentences increasing in length and complexity • Basic Concepts sub-test: measures understanding of different semantic concepts, including: dimension/size, direction/location/position, number/quantity. • The Concepts and Following Directions sub-test: tests children's ability to respond to instructions of increasing length/complexity, as well as particular characteristics of objects (e.g. size, ordinal position)
Auditory Comprehension sub-scale - Preschool Language Scales - V	Zimmerman, Pond & Steiner (2011)	Birth – 7:11 years	unknown	Understanding of a wide range of semantic concepts (e.g. colours, objects, spatial terms) as well as the ability to draw inferences

3.4.5.3 Approaches to learning

In order to assess AtL, teacher ratings of children's behaviour towards learning tasks in the classroom were used, given that teachers are well-positioned to observe children's behaviour in a wide variety of different natural learning contexts over time.

A range of different teacher-rating scales that measure aspects of AtL in young children were identified from current literature (Table 3.7). Following inspection of items across different scales, the Initiative sub-scale of the Devereux Early Childhood Assessment for Preschoolers, 2nd edition (DECA-P2) (LeBuffe & Naglieri¹⁴, 2012), was thought to most closely assess those aspects of AtL hypothesised to be related to the *onebillion* intervention, such as independence, persistence and confidence.

The Initiative sub-scale of the DECA-P2 is part of a larger teacher questionnaire (38 items), which also measures other dimensions of children's social-emotional development. Teachers are asked to rate how often children demonstrate different behaviours on a 5-point Likert scale from "Never" to "Very Frequently". The internal consistency of this scale is high (Cronbach's alpha = .92) and test-retest reliability is good ($r = .89$). Previous research has also demonstrated that ratings on this scale predict concurrent mathematics attainment (Dobbs, Doctoroff, Fisher, & Arnold, 2006).

With the permission of the authors, the researcher constructed a shorter 9-item teacher questionnaire consisting only of those items constituting the 'Initiative' scale. In the original version of the questionnaire, teachers were also asked to rate children's behaviour based on their observations in the past four weeks. For the present research, the period for observation was shortened to two weeks in order to be more sensitive to changes over the nine week intervention period. A total score was calculated on each item and this score was used in analyses.

¹⁴ with the Devereux Center for Resilient Children

Table 3.7: Approach to learning teacher rating scales for children aged 3-4 years

Name of Measure	Source	Age Range	Scale	Type of social-emotional skills
DECA-P2	LeBuffe & Naglieri (2012) <i>Devereux Center for Resilient Children</i>	3:00-5:00 years	38 items Likert 1-5	Tests children's initiative (ability to use independent thought and action to meet their needs), self-regulation (ability to express and manage emotions), attachment and relationships (ability to promote positive relationships with others)
Preschool Learning Behaviours Scale	McDermott, Green, Francis & Stott (2000)	3:00-5:06 years	27 items Likert 1-3	Tests three different dimensions of learning behaviour, including: competence and motivation to learn, attention and persistence with learning and attitude toward learning
Learning to Learn Scales	McDermott et al. (2011)	3:02-5:09 years	55 items Likert 1-3	Tests seven dimensions: strategic planning, effectiveness motivation (perseverance), interpersonal responsiveness in learning, vocal engagement in learning, sustained focus in learning, acceptance of novelty and risk and group learning
Child Behaviour Rating Scale – <i>Two versions</i>	Bronson, Goodson, Layzer & Love (1990)	3:00-6:00 years	27 items (Version 1) 17 items (Version 2) Likert 1-5	The original version consists of: a mastery behaviours scale (assessing children's self-regulation and engagement) and a separate social skills scale An updated version obtained from the developers consists of two refined scales: self-regulation and interpersonal skills

3.4.5.4 Administration of measures

Direct child assessment measures were administered individually by the researcher in a quiet room or area at the nursery. Testing consisted of three short sessions with each child, each lasting no longer than 15 minutes. All tests were administered in the same order and according to standardised instructions. Administration of Form A/B of the mathematics CK measure was counterbalanced across participants and time.

The main class teacher of each nursery child was asked to complete the AtL questionnaires over the course of a week (following standardised instructions). At school A, the teacher did not return questionnaires for the control group until after the six week summer holiday. Possible threats to reliability due to this delay are discussed further in Section 3.4.7.

3.4.6 Procedure

3.4.6.1 T1: Pre-test measures and group allocation

Following administration of pre-test measures, children were allocated to experimental and control conditions. In order to establish equivalence in pre-test ability across groups, children were allocated to condition according to a ranking procedure. Within each class, children were ordered from highest to lowest according to their ability scores on the ENC sub-test. This measure was selected for matching due to its established high reliability and validity (see Section 3.4.5). Children were then alternately placed in either the intervention or control conditions as follows:

Morning class at Nursery A: 12121212...

Afternoon class at Nursery A: 21212121...

Morning class at Nursery B: 12211221...

Children who did not participate in individual assessments were randomly allocated to experimental or control conditions.

3.4.6.2 T1-T2: Intervention period

Following pre-testing, children in the experimental condition received the intervention for 15 minute sessions on a daily basis over the course of 9 weeks, whilst children in the control condition accessed self-directed play activities.

Children accessed the intervention in small groups of between 4-8 children. Sessions were facilitated by nursery class teachers or support staff, as shown in Table 3.8. Further information about the facilitators is provided in Section 3.5.2. Staff were asked to deliver the intervention at a regular time each day, allowing some flexibility to take into account wider school events (e.g. play rehearsals).

Table 3.8: Delivery of the onebillion intervention for the experimental condition

Nursery	Session	Number of Children in Group	Facilitator
Nursery A	Morning	8	Class teacher - Anne (All sessions)
	Afternoon	5	Class teacher – Anne (Tue – Fri) Teaching assistant – Bianca (Mon)
Nursery B	Morning	10 children, split into two groups of 5	Class teacher – Clara (Wed – Fri) Teaching assistant – Danielle (Mon - Tue)

Prior to implementation of the intervention, all facilitators received training by the researcher at the school setting (see Appendix 8.10). Facilitators were supported to download the app software onto the school iPad tablets and trained in how to support the children whilst they were using the apps. They were also asked to record children's progress through topics on the app using a tracking grid.

3.4.6.3 T2 and T3: Post-test measures and longitudinal follow-up

Immediately following the intervention period (T2), post-test measures of mathematics, receptive language skills and AtL were repeated for all children in the experimental and control conditions.

Measures of mathematics were administered again 5 months after the end of the intervention period (T3) for children in both conditions to determine whether learning gains were sustained over time. In the period between T2 and T3, none of the children in the project accessed the *onebillion* intervention at school.

3.4.7 Validity and reliability

In line with the philosophical assumptions of post-positivist research, this section considers potential threats to the validity and reliability of the evaluation research design and the steps taken by the researcher to address each threat. Limitations to the design are explicitly acknowledged.

3.4.7.1 Internal validity

Internal validity refers to the extent to which a particular experimental design is able to demonstrate a causal relationship between the manipulation of an IV and changes in a DV (Mertens, 2015). Table 3.9 outlines potential threats to internal validity which need to be considered in experimental research (Campbell & Stanley, 1963; Cook & Campbell, 1979) and indicates how each was addressed in the present research.

Table 3.9: Internal validity of the research design

Threat	Definition of Threat	Action taken to address threat
History	Changes in the external environment unrelated to the introduction of the intervention may affect the DV(s)	Inclusion of a control group from the same class
Testing	Changes in performance may occur as a result of repeated testing	<ul style="list-style-type: none"> • A gap of at least 9 weeks between repeated uses of experimental measures • Two parallel forms of the experimenter-designed maths measure were used at T1 and T2 (order counterbalanced)
Instrumentation	Changes in measurement at pre and post-test	Standardised delivery of assessments by the same assessor at each time point. Some limitations are acknowledged due to the researcher's lack of familiarity with assessments and a delay in the return of teacher questionnaires at Nursery A (see Table 3.10).
Regression	Extreme groups score closer to the mean at post-test	An even spread of mathematics scores at pre-test across control and experimental groups was achieved through ability-matching
Mortality	Attrition of participants	Registers were taken to monitor attendance at sessions. Attrition from the study was low overall: <ul style="list-style-type: none"> • Parental consent for one participant was withdrawn by T2. • Three participants had moved schools by T3 and were unavailable for reassessment.
Maturation	Natural development and growth in participants over time	Inclusion of a control group from the same class

Threat	Definition of Threat	Action taken to address threat
Selection	Initial systematic differences between participants in each group e.g. ability, background	Selection of control and experimental groups at the pupil level based upon pre-test mathematics ability scores rather than use of pre-existing group.
Diffusion of treatments	Control group receives aspects of intervention	Clear records were kept for teaching staff regarding pupil allocation to groups. Nursery staff confirmed that children in the control condition did not access the <i>onebillion</i> intervention during nursery sessions.
Compensatory equalisation of treatments	Organisational pressures for the control group to receive equal benefits may lead them to be provided with additional resources in the interest of fairness	Teachers confirmed that pupils continued to receive normal mathematics teaching and all children had access to iPad tablets for other activities during the intervention period. However, as children in both groups were taught by the same class teacher, this is acknowledged as a limitation of the present research.
Compensatory rivalry	Children in the control group might be motivated to improve their performance to compete with the experimental group	Due to the age of the children, this was not likely to be a high threat in the present study. However, it is acknowledged that parents may have provided additional instruction to children in the control group.

3.4.7.2 *External validity*

External validity refers to the extent to which the research findings are applicable beyond the particular context of the research study (Mertens, 2015). This question is particularly important for educators and policy-makers in understanding the applicability of research to new situations and addressing the question ‘Will it work here?’ (Green et al., 2015). Despite the small scale of the present research, a number of steps were taken by the researcher to enhance the external validity of the current research. First, a clear description of the nature of the intervention and its implementation is outlined here to ensure that these procedures can be replicated. Second, demographic characteristics of participating school-based nurseries and children are stated, highlighting the population within which the findings are most applicable. Finally, through the use of a mixed methods design, the possible relationships between intervention, population and outcome have been explored to determine the contextual factors that may have influenced the outcomes of the intervention (Green et al., 2015).

However, limitations to the external validity of the design are also acknowledged. First, it is possible that any significant effects could be due to the experimental group receiving additional special attention from facilitators, known as a Hawthorne effect (Roethlisberger & Dickson, 1939). Second, it is possible that the intervention may have an effect only because it is novel, or alternatively, may fail to have an effect because it causes some disruption to typical activities when first introduced (Mertens, 2015).

3.4.7.3 *Additional threats to validity*

3.4.7.3.1 *Intervention fidelity*

In order to monitor the fidelity of the intervention, facilitators recorded any days when it was not possible to deliver the intervention and completed an attendance register. Out of 44 possible intervention sessions over the 9 week period (excluding a bank holiday), 40 sessions were delivered at Nursery A and 41 sessions at Nursery B. Registers showed that the number of sessions actually received by children ranged from 31 to 41 sessions due to individual absences.

The researcher also conducted three checks of intervention fidelity per group, distributed over the course of the intervention. During these checks, the researcher observed whether the intervention was implemented as specified during training sessions (see Appendix 8.11). As displayed in Table 3.10, intervention fidelity was typically high. At Nursery A, however, there was a technical problem with the headphone ports on the iPad tablets, which meant that two children in each group were unable to use headphones during the intervention. These two children were subsequently seated separately from each other to minimise noise disturbance.

Table 3.10: Intervention fidelity checks

Feature of the Intervention	% of sessions observed
Intervention lasted for 15 minutes	92%
Children accessed their own profile within the app within the first 3 minutes of the intervention	100%
Children progressed in sequence through activities on the app	92%
Children wore headphones whilst using the app	58%

3.4.7.3.2 *Length of treatment*

It is important that the duration and frequency of any intervention is sufficient to ensure that any treatment effect can be observed (Mertens, 2015). A previous evaluation of the *onebillion* intervention found a moderate to large intervention effect after 8 weeks (Outhwaite et al., 2017). Implementing daily intervention sessions over a period of 9 weeks should therefore have been sufficient to establish whether the intervention had a significant effect.

3.4.7.4 *Reliability*

A number of different potential threats to reliability in experimental research are identified and discussed by Robson (2011). In the present research, steps were taken to address each of these threats, as detailed in Table 3.11.

Table 3.11: Threats to reliability

Threat	Definition of Threat	Action taken
Participant error	Changes in the performance of participants due to factors unrelated to experimental manipulation (e.g. tiredness)	<ul style="list-style-type: none"> Measures were administered in a standardised order. Assessments were reorganised if teacher had concerns about the child's affect or tiredness.
Participant bias	Participants show enhanced performance to meet the aims of the intervention, or reduced performance due to disaffection	Preschool children were unaware of the full aims of the research due to their age.
Observer error	Observer records responses incorrectly	<ul style="list-style-type: none"> The researcher aimed to administer all tests according to standardised instructions. Teachers were asked to complete AtL questionnaires in a quiet area Teachers were asked to complete all questionnaires within a week based on observations of children's behaviour in the preceding two weeks.
Observer bias	Observer is biased to demonstrate the effectiveness of the intervention (perhaps for reasons of investment in time delivering the intervention), or, alternatively, a lack of effect (perhaps due to disaffection)	<p>Observer bias is acknowledged as a limitation.</p> <ul style="list-style-type: none"> It was not possible for the researcher administering measures to be blind to condition due to the need for fidelity checks and observations to be conducted by the same researcher. Teachers completing post-test AtL questionnaires were also aware of participant's condition.

There were two notable threats, however, to the reliability of the findings in this study which should be acknowledged. First, there was a delay in the teacher's return of questionnaires for children in the control group at Nursery A until after a 6 week summer holiday period following the end of the intervention. Teacher recall of children's behaviour may therefore have affected the reliability of these results.

Second, the researcher also made an error in the administration of the BAS-III ENC sub-test for two children (one at pre-test and one at post-test), stopping the test too soon due to a misinterpretation of the standardised instructions in the administration guidance for the test (Elliot & Smith, 2011). Incorrect use of stopping points may have affected the reliability of these assessment scores; however, as removal of these scores from the analyses did not affect the pattern of the results or the statistical significance of findings, data analyses presented later (Section 4.1.2) are for the full sample.

3.5 Embedded Aspects of the Research Design

3.5.1 Research Questions 5 and 6

Additional explanatory data were gathered to further illuminate the outcomes of the study and the mechanisms which may be underpinning the quasi-experimental outcomes, addressing the following research questions:

5.) What are facilitators' perceptions of the outcomes of the onebillion intervention for preschool children aged 3-4 years?

6.) What factors may affect the outcomes of the onebillion intervention for preschool children aged 3-4 years?

Methods used for the sampling and data collection within embedded aspects of the design are discussed below.

3.5.2 Semi-structured facilitator interviews

3.5.2.1 Rationale for conducting semi-structured interviews

In order to answer Research Questions 5 and 6, data were collected about facilitators' perceptions of intervention outcomes and the factors which they felt

may have affected these outcomes. Whilst questionnaires or focus groups were considered, semi-structured interviews were selected as the most appropriate method to collect detailed information about each individual's perspective.

During semi-structured interviews, the researcher uses an interview schedule to guide questioning and ensure that research questions are addressed. Some flexibility is also afforded, however, as the researcher can alter the wording of questions and probe participants' responses further (Weiss, 1994). The advantage of this approach is that it preserves the natural flow of conversation and the researcher can follow-up aspects of the discourse that they consider important (Coolican, 2014). Nevertheless, it is acknowledged that comparison across participants and the dependability of the research may be weakened by a lack of standardisation across interviews (Coolican, 2014; Robson, 2011).

3.5.2.2 Facilitator selection

The researcher met with each facilitator to share information about the nature of the study, gaining written informed consent from all four facilitators to participate in interviews (Appendix 8.12). Facilitators provided key details about their experience and current attitude towards the use of technology at the start of each interview, as displayed in Table 3.12, given that these factors might influence their views and experiences implementing the intervention (see Shanley et al., 2007).

3.5.2.3 Interview procedure

The researcher conducted the interviews individually in a quiet room at each school, following the schedule in Appendix 8.13. All interviews were recorded and lasted between 20-45 minutes. Participants were debriefed at the end of the interview (see Appendix 8.14).

Table 3.12: Facilitator details, including confidence, attitude and experience in using technology¹⁵

Pseudonym	Job Title	Length of Experience in Working with Early Years Children <i>Supporting Qualifications</i>	Uses technology for personal use Yes (Y) or No (N)	Previous experience using apps with children Yes (Y) or No (N)	Confidence using Educational Technology¹⁶ <i>Likert scale 0 (low) to 5 (high)</i>	Perceived Value of Educational Technology¹⁷ <i>Likert scale 0 (low) to 5 (high)</i>
Anne <i>Nursery A</i>	Class Teacher	6-7 years <i>Qualified Teacher</i>	N	Y	4	4
Bianca <i>Nursery A</i>	Higher-level Teaching Assistant (HLTA)	Approximately 19 years <i>NVQ Level 3 in Early Years HLTA qualification</i>	Y	N	4	4/5
Clara <i>Nursery B</i>	Class Teacher	Approximately 15 years <i>Qualified Teacher</i>	Y	Y	3	5
Danielle <i>Nursery B</i>	Nursery Nurse/ HLTA	16 years <i>NVQ Level 3 in Early Years HLTA qualification</i>	Y	Y	4	4

¹⁵ Facilitator attitude/experience may affect implementation of technology-based interventions (Shanley et al., 2007) and are therefore reported here.

¹⁶ Facilitator's response to the question, "How confident are you in using educational technology within early years education?"

¹⁷ Facilitator's response to the question, "How beneficial do you feel that educational technology can be within early years education?"

3.5.3 Narrative observations

In order to further explore the factors that may have affected outcomes, the researcher also conducted narrative field observations whilst the children were using the apps. The purpose of these observations was to: (a) triangulate the data obtained through facilitator interviews and (b) provide a greater understanding of children's own experiences using the intervention.

One to two narrative observations were conducted per group, ensuring that each facilitator was observed at least once with each group (*7 observations in total*). The researcher took an exploratory stance to these observations, noting any salient factors that may have been affecting outcomes. However, a number of key foci were identified, including:

- The support given by the facilitator
- Children's progression through the activities and use of the app software
- Children's response to the app software
- Implementation of the intervention in the setting

The researcher aimed to distribute her attention evenly across all children during the intervention sessions. Nevertheless, it is likely that the researcher's own prior interests, experience and expectations affected her attention to particular events, and her encoding/interpretation of those events (McCall, 1984), as discussed further in Section 4.2.1.

In addition, there may have been some reactive effects, where the presence of the researcher influenced the behaviour of participants to some extent (McCall, 1984; Robson, 2011). The researcher took steps to minimise reactivity by explicitly asking facilitators to conduct the group as normal and not to interact with the observer during the session. The children had already gained familiarity with the researcher prior to the sessions during individual assessments and therefore typically accepted the researcher's presence without seeking interaction.

3.5.4 Structured observations of attention on-task

Structured observations of the intervention sessions were also conducted in order to explore whether children's attention on-task and engagement in learning may have affected intervention outcomes. Three observations were conducted by the researcher per group, distributed across the intervention period and ensuring that each facilitator was observed at least once.

An observation schedule was adapted from the work of Merrett and Wheldall (1986), who established a protocol for observing the on and off-task behaviour of groups of pupils in the classroom. Observations lasted 10 minutes, commencing 3 minutes after the session began. Each child in the group was observed in turn to ascertain whether they were on-task, off-task or they had temporarily left the group. Using a time sampling schedule, children were observed for 5 seconds, followed by a 5 second non-observing interval when the researcher's judgement was recorded. In order to eliminate bias, the first child in the group to be observed during each session was determined randomly by rolling a dice.

On-task behaviour was defined as the child doing any of the following actions throughout the whole observation interval:

- Looking at the iPad tablet screen
- Using the app software
- Receiving facilitator support
- Receiving or giving peer support

In total, 60 observation intervals were recorded across each intervention session, distributed evenly across each child in the group. The percentage of on and off-task behaviours were then calculated for the group as a whole for each intervention session. This procedure supported descriptive analysis of children's engagement over time and across different settings.

In order to reduce observer error and increase reliability, the non-observer interval was lengthened from 1 second to 5 seconds following pilot testing. However, it is important to acknowledge the possibility of observer bias and

error in the interpretation and encoding of events, as well as observer effects on the behaviour of participants (McCall, 1984).

3.5.5 Associations between learner characteristics and intervention gains

Finally, further quantitative analyses were conducted in order to determine whether any particular characteristics of the children may have affected the extent to which they benefited from the intervention (Research Question 6). Correlational analyses were used to identify any significant associations between children's learning gains and their AtL receptive language skills, maths attainment and SES at the start of the intervention. Assessments used to measure each DV are described in Section 3.4.5. SES was measured through the calculation of an IDACI score¹⁸ based on each child's postcode (see Section 3.5.3).

¹⁸ Instead of IDACI ranks, reported in Section 3.4.3, more precise IDACI scores, ranging from 0 to 1, were used in this analysis; higher scores indicate greater levels of deprivation.

3.6 Ethical Considerations

The British Psychological Society (BPS, 2014) emphasises that researchers have a responsibility to conduct ethical research which establishes “mutual trust and confidence between investigators and participants” and which respects “the rights and dignity of participants in their research and the legitimate interests of stakeholders...and society at large” (p.4). Throughout the design and implementation of this research, the researcher took a number of steps to ensure that the study was ethical, making reference to guidance available from the following sources:

- Code of Human Research Ethics (BPS, 2014)
- Code of Ethics and Conduct (BPS, 2009)
- Standards of Conduct, Performance and Ethics (Health Care & Professions Council, HCPC, 2012)
- Code of Research Conduct and Research Ethics (UoN, 2015)

Prior to conducting the study, the researcher received ethical approval from the UoN Research Ethics Committee (see Appendix 8.15). The key ethical considerations addressed in the current research are discussed in Table 3.13.

Table 3.13: Ethical considerations applicable to the present research

Ethical Consideration	Details of Concern	Action Taken
Informed Consent	<p>Written informed consent should be obtained from all research participants. This will include ensuring that participants understand the nature, purpose and possible consequences of their involvement (BPS, 2014; BPS, 2009 – Standard 1.3; HCPC, 2012 – Standard 9; UoN, 2015).</p> <p>Consent of participants regardless of age and competence level should be sought. For children under the age of 16 years additional consent should be gained from their parents/carers. (BPS, 2014).</p>	<ul style="list-style-type: none"> • Signed informed consent was obtained from the head teacher, parents and facilitators (see Appendices 8.6, 8.7, 8.12). • Children provided assent to participate in assessments. The researcher explained the project to the children using a pre-developed script (Appendix 8.8) and gave opportunity for questions. Children were asked to indicate their willingness to participate by responding verbally or pointing at a smiley face (agreement) or a frowning face (wishing to return to class). This occurred on two occasions and children were allowed to withdraw from the research. Responses suggested that children did not feel coerced into participating.

Ethical Consideration	Details of Concern	Action Taken
Right to Withdraw	<p>All participants should be aware of their right to withdraw from the research and not answer particular questions (BPS, 2014; BPS, 2009 – Standard 1.4).</p> <p>In the case of very young children, continued assent to participate should be carefully monitored by attention to verbal/non-verbal signs that they are not willing to continue (BPS, 2014).</p>	<ul style="list-style-type: none"> • The researcher asked children for renewed assent prior to each assessment session. • Children were told that they did not have to answer any questions and that they could return to class at any time. Due to the age of participants, the researcher also monitored any verbal/non-verbal signs during assessments that children did not wish to continue. • School staff and parents were informed of their right to withdraw at any time, without giving a reason, and that this would not affect their access to other support from the EPS. Facilitators were informed that their withdrawal from the interviews would not affect wider school participation.
Confidentiality	<p>Psychologists and researchers should respect individual's right to privacy and confidentiality (BPS, 2014; BPS, 2009 – Standard 1.2; HCPC, 2012 – Standard 2).</p> <p>All data should be collected, process and stored in accordance with the Data Protection Act to prevent inadvertent disclosure (BPS, 2014; UoN, 2015).</p> <p>Data should be anonymised so that individuals are not personally identifiable (BPS, 2014).</p>	<ul style="list-style-type: none"> • All data has been fully anonymised in the reporting of this research. • Consent forms and data has been stored securely in password protected files/locked filing cabinet and only accessed by the researcher. • Confidentiality of child participants was ensured by assigning each individual a unique identifier to record assessment data anonymously and to track data collected over time. • Schools were provided with non-anonymised assessment data to inform teaching practice. Parents were able to access data about their own child on request from the school. • The purpose of collecting postcode data as a measure of SES was explicitly stated on consent forms and parents were given the option not to disclose their postcode even if they consented.

Ethical Consideration	Details of Concern	Action Taken
Minimising harm	The research should take steps to minimise any potential for risks to participants' psychological well-being, health, values or dignity (BPS, 2014; BPS, 2009 - Standard 3.3; HCPC, 2012 – Standard 1).	<ul style="list-style-type: none"> • The researcher administering the tests held a full Disclosure and Barring Service check and had professional experience working with young children. • During assessment sessions, time was taken to build rapport with each child. All sessions typically took place in a room or quiet area close to the child's classroom and a familiar adult was available. • All children continued to access their typical mathematics teaching. • All children had regular access to iPad tablets at different points in the day to reduce the possibility of conflict between groups. • Nursery staff were given opportunity to review the interview transcripts prior to analysis and to ask for any data to be removed. These checks aimed to ensure that staff did not feel any regret for the answers that they gave. All staff consented to their full transcripts being used in the analysis. • Any children experiencing particular distress during the intervention were removed from the study; this occurred after a parent raised concerns that their child had been upset about wearing headphones and using the app. Nursery staff were asked to provide any children experiencing frustration or not making progress with additional teaching support. • The <i>onebillion</i> app software was provided free of charge to schools in order to ensure that they would not be financially disadvantaged through participation in the research.

Ethical Consideration	Details of Concern	Action Taken
Debriefing	<p>All participants should be debriefed at the end of research to inform them of the outcomes, identify unforeseen harm/misconceptions and provide support if needed (BPS, 2014; BPS, 2009 - Standard 3.4; HCPC, 2012 – Standard 7).</p>	<ul style="list-style-type: none"> • Debrief statements were read aloud to children at the end of individual assessments and to nursery staff following the interviews (Appendix 8.14). • A summary of research findings was given to school staff and parents.
Scientific integrity and social responsibility	<p>Research should be carefully designed and conducted to ensure quality and integrity. The aim of research should be to contribute to the benefit of society and researchers should not cause unnecessary disruption to the social context in which they work (BPS, 2014; BPS, 2009 - Standard 4.1; HCPC, 2012 – Standard 13; UoN, 2015).</p> <p>Researchers should seek to maximise the benefit of their research at all stages, including dissemination (BPS, 2014).</p>	<ul style="list-style-type: none"> • The author sought supervision regularly and discussed the study with other researchers who had conducted previous evaluations of this intervention. • The unique contribution of this research to the evidence-base around tablet-based interventions has been explicitly stated. • Care was taken to set up this research in collaboration with teaching staff to minimise disruption. • A summary of research findings was given to key stakeholders, including the organisation <i>onebillion</i>, schools and parents. • The author has no financial links with the <i>onebillion</i> intervention, other than their agreement to supply the app software free of charge to participating schools. The research design, analysis, interpretation and reporting of findings were conducted independently from the organisation. • The researcher acknowledges that she has a relationship with the UoN research team that has previously found positive outcomes from use of the <i>onebillion</i> intervention; however, the researcher strived to interpret the outcomes of this study independently from past studies in order to ensure a fair evaluation of use of the apps with younger children.

3.7 Summary

This chapter began by discussing the philosophical positioning of the current study within the pragmatic research paradigm. A rationale was then provided for the mixed methods embedded design selected to evaluate the *onebillion* intervention, followed by a presentation of the specific sampling methods, measure and procedures that were used. Steps taken to address issues of validity, reliability and ethics were also considered. In the next chapter, procedures used for data analysis are discussed and the results of this study are presented.

4 Results

This chapter presents the results of the study in relation to each of the research questions stated in Section 2.9. Findings are presented in two sections, beginning with the quantitative data analyses used to evaluate outcomes from the quasi-experimental evaluation of the *onebillion* intervention. Data analyses from embedded aspects of the research design, including qualitative interviews and observations, are then discussed.

4.1 Quasi-Experimental Evaluation Data Analysis

Quantitative data analyses of the quasi-experimental design focused on addressing research questions 1-4, evaluating the impact of the intervention on key DVs (mathematics, receptive language and AtL) at post-test and longitudinal follow up. A rationale for the statistical analysis procedures used in this research are discussed in the following sections, followed by the results of these analyses.

4.1.1 Data analysis procedures

4.1.1.1 *Descriptive statistics*

Within the present study, mean averages and standard deviations (SDs) are reported to give a measure of the central tendency and spread of the data respectively. Both of these statistics, however, can be distorted by extreme values and skewed data (Bryman & Cramer, 2011). The distribution of all relevant data were therefore analysed (see Appendix 8.16). Where data are not normally distributed (displaying symmetry about the mean), the median and interquartile range have been reported.

4.1.1.2 *Inferential statistics*

Inferential statistical tests have been also been used in this research in order to test the hypotheses stated in Section 3.4.1 for each research question. Where the probability value, p , given by the test is $< .05$, it is typically accepted that there is a *statistically significant* finding and the researcher can be “reasonably confident” in rejecting the null hypothesis (Coolican, 2014); at this significance

level, there is only a 5% chance of the researcher committing a Type I error, rejecting the null hypothesis when it is actually true.

4.1.1.3 Parametric and non-parametric tests

There are two types of inferential statistical analysis: parametric and non-parametric. Parametric tests are usually preferable as they have greater statistical power than their non-parametric alternatives (Coolican, 2014). They also allow more complex analyses to be conducted, such as controlling for the effects of a confounding variable. During parametric testing, the following statistical assumptions are made about the data:

1. Data must be at an interval of ratio level (i.e. continuous).
2. The sample should be taken from a population which is normally distributed.
3. The samples being compared must be drawn from populations with the same variance (otherwise known as homogeneity of variance).

(Brace, Kemp & Snelgar 2012, p. 12)

In order to determine whether the data met these criteria a number of preliminary checks were conducted (see Appendix 8.16). Where any assumptions were violated, the researcher used non-parametric testing instead.

4.1.1.4 Choice of parametric statistical analysis

Two different types of parametric statistical analysis can be used to evaluate the impact of an intervention on DVs: analysis of gain scores or analysis of covariance (ANCOVA) (van Breukelen, 2013). In the first model, gain scores are calculated as the difference between pre- and post-test scores on a particular DV. A t-test can then be used to determine whether there were significant differences between experimental conditions in learning gains during the intervention period. A key limitation of gain score analysis, however, is that the effect of initial attainment on outcomes may not be fully removed from the analysis, given that they are likely to correlate with learning gains (Dugard & Todman, 1995), see Section 3.4.2.

In contrast, ANCOVA allows the researcher to test whether there are any differences between conditions on a particular DV, whilst statistically controlling for the possible effects of a confounding variable (covariate) (Pallant, 2016). In

an intervention study, experimental and control groups' performance on particular outcome measures can therefore be compared, whilst adjusting for covariance in children's scores at pre-test (van Breukelen, 2013). Despite the matching procedures used in the present study, ANCOVAs were selected as the most appropriate form of analysis as they would control for any residual differences between groups on baseline skills.

In addition to meeting assumptions for parametric testing, the researcher also checked whether data met two other statistical requirements which are assumed in ANCOVA (see Appendix 8.16):

- 1.) There should be a linear relationship between the DV and covariate.
- 2.) There should be homogeneity of regression across all experimental groups i.e. a similar relationship between the DV and covariate across all conditions.

(Brace et al., 2012)

4.1.1.5 *Effect size*

Whilst inferential statistics indicate the *statistical* significance of an effect, they have limited value in evaluating the *practical* significance of findings as the statistical power of inferential analysis to identify treatment effects is always affected by sample size (Clark-Carter, 2007). Where the sample size is large, the *p* value of the test might indicate a statistically significant result, even when the actual magnitude of the difference between intervention and control groups is only small. Conversely, when the sample size is small, the statistical power of the test to detect a significant effect may be limited, even where there is a large group difference.

ES is a measure of the magnitude of an experimental effect without conflating sample size (Clark-Carter, 2007). It allows the researcher to "move beyond the simplistic 'Did it work (or not)?' to the far more important 'How *well* did it work?'" (Higgins, Kokotsaki, & Coe, 2012, p.7). Measures of ES are therefore particularly valuable for educators implementing interventions in schools and policy-makers deciding how best to allocate funding and resources. One of the most common and standardised measures of ES is Cohen's *d*, where:

$$d = \frac{\text{difference in group means}}{\text{pooled standard deviation}}$$

In general, a Cohen's *d* value of 0.2 is considered *small*, 0.5 is *moderate* and 0.8 is *large* (Cohen, 1988). These labels are, however, somewhat arbitrary and values should be considered in context (Cohen, 1988); it has been argued, for example, that even an ES as small as 0.1 could be educationally significant, particularly if the intervention was easy to implement, inexpensive and led to cumulative benefits over time (Higgins, Kokotsaki et al., 2012).

4.1.1.6 Statistical power

As noted above, statistical power is affected by ES, variance, the number of participants, and the significance level of the test (Brace et al., 2012). Where statistical power is too low, typically less than 80%, the researcher runs a greater risk of committing a Type II error, failing to reject the null hypothesis when it is false (Nuzzo, 2016).

A sample size calculator, G Power 3.1, was used to determine the sample size that would be needed to achieve an 80% power level in the current study. Assuming a large ES, as found in previous research with low attaining 4-5 year old children (Outhwaite et al., in press), an overall sample size of 52 would be required, with 26 participants in each group. It is recognised that the sample sizes included in the analyses fall slightly below these requirements (maximum $n = 47$). Therefore, there is potential that the effects of treatment may be missed, particularly if the magnitude of effects are smaller. Consideration of ES alongside significance testing is therefore particularly important in the present study.

The results of these quantitative analyses are presented in the following sections in relation to each research question.

4.1.2 Data analysis for Research Question 1

The study aimed to address the following primary research question:

What is the impact of the onebillion tablet intervention on the mathematical attainment of preschool children aged 3-4 years?

The DVs were related to children's mathematical attainment, as measured by their ability scores on the ENC maths sub-test (Elliot & Smith, 2011) and their scores on the researcher-developed CK assessment (Outhwaite et al., 2017). It was hypothesised that at post-test (T2) children in the intervention group would have significantly higher attainment in mathematics on both measures, compared to children in the control group. Analyses for each measure of attainment are presented separately below, beginning with those conducted from children's ENC ability scores.

4.1.2.1 Early Number Concepts: Descriptive statistics

A full set of data was obtained at T1 and T2 for all 43 children who consented and participated in individual assessments. Descriptive statistics for the ENC data gathered are displayed in Table 4.1 and Figure 4.1, with scores provided separately by nursery.

Table 4.1: Descriptive statistics for Early Number Concepts ability scores across experimental and control conditions at T1 and T2

Condition	Sample	Pre-test Mean Ability Score (SD)	Post-test Mean Ability Score (SD)	Mean Learning Gains (SD)
Control	Nursery A (n = 13)	96.23 (21.56)	107.00 (25.26)	10.77 (13.55)
	Nursery B (n = 9)	85.89 (22.68)	95.78 (24.08)	9.89 (17.05)
	Overall (n = 22)	92.00 (22.10)	102.41 (24.85)	10.41 (14.69)
Experimental	Nursery A (n = 13)	96.08 (21.41)	107.62 (22.49)	11.54 (9.13)
	Nursery B (n = 8)	87.50 (18.34)	108.50 (15.20)	21.00 (17.92)
	Overall (n = 21)	92.81 (20.28)	107.95 (19.61)	15.14 (13.58)

Figure 4.1 indicates that there was minimal difference between the overall mean ENC ability scores in the control and experimental conditions at T1. Whilst the mean scores increased for both conditions between T1 and T2, this increase was larger for the experimental group.

Separate analysis by nursery, however, indicates a different patterns across settings. At Nursery A, mean scores for the control and experimental groups were similar to each other at both time points. In contrast, at Nursery B, where ENC ability was initially lower, on average, children in the experimental group made larger gains than those in the control group, closing the gap in initial attainment with children in Nursery A by T2.

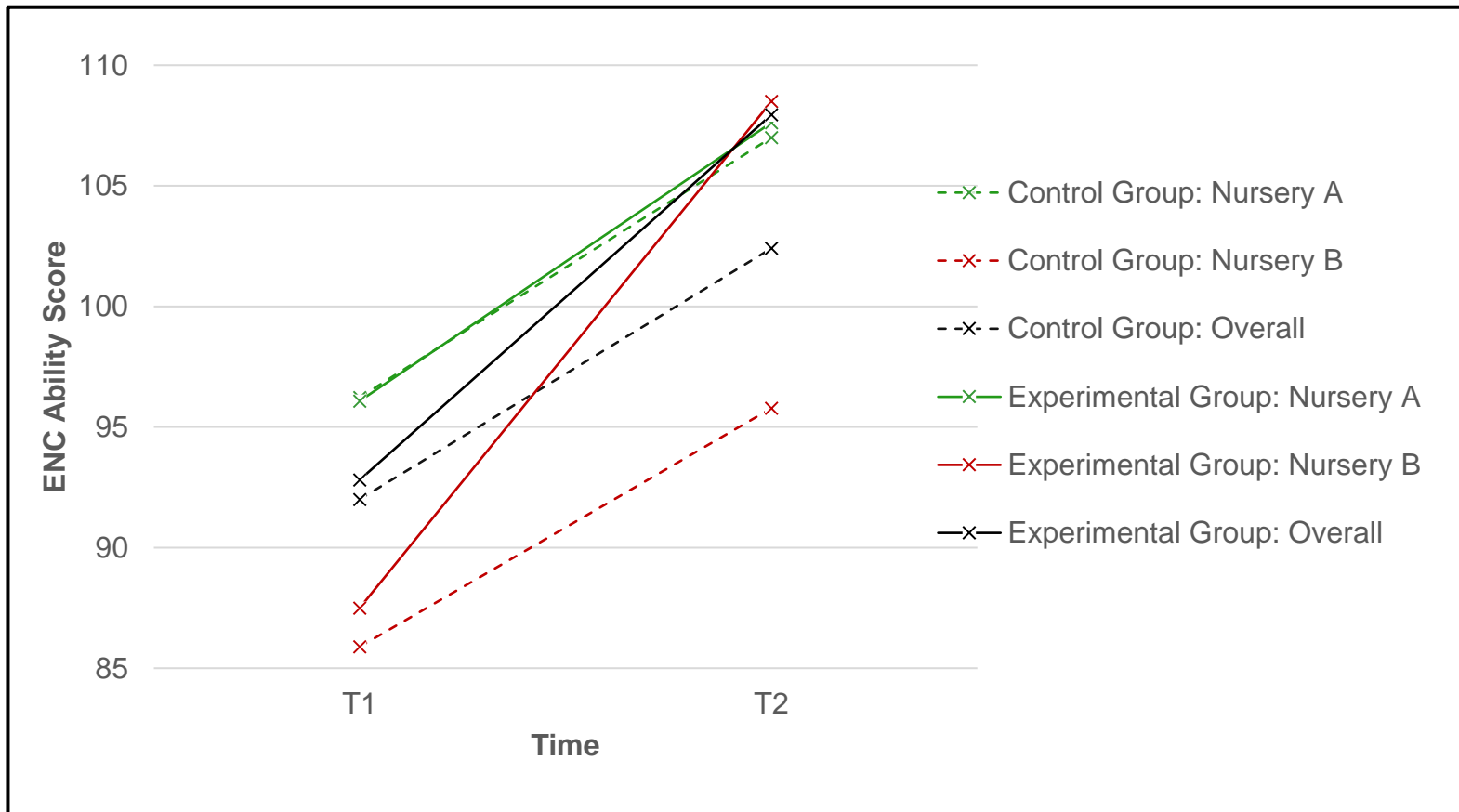


Figure 4.1: A line graph to show change in experimental and control conditions over time in Early Number Concept ability scores

4.1.2.2 Early Number Concepts: Inferential statistics

ANCOVA was used to determine whether there was a statistically significant difference between ENC Ability Scores at post-test (DV) between experimental conditions, after controlling for pre-test scores (covariate). Preliminary checks indicated that all necessary statistical assumptions of ANCOVA held for this data set (see Appendix 8.16). The results indicated that there was not a statistically significant difference between the experimental and control groups at post-test on the ENC assessment, after adjusting for pre-test scores, $F(1, 40) = .13$, $p = .258$, $\eta^2 = .032$.

A two-tailed independent samples t-test indicated that there was no statistically significant difference between groups on T1 ENC ability scores, $t(41) = 1.25$, $p = .901$, indicating that matching had been successful.

4.1.2.3 Early Number Concepts: Effect size

Although the ANCOVA results were not statistically significant, a possible trend was indicated by the descriptive statistics. Given low statistical power, an ES analysis, using Cohen's d was therefore conducted to determine the practical significance of the difference between experimental conditions at post-test, see Table 4.2, suggesting that overall the intervention had a 'small' effect on attainment on this measure.

Table 4.2: Effect size analysis of the difference between conditions in T2 Early Number Concepts ability scores

	Unadjusted means (SD)	Cohen's d	Size of Effect
Total Sample ($n = 43$)	Control: 102.41 (24.85) Experimental: 107.95 (19.61)	+ 0.25	Small

4.1.2.4 Curriculum Knowledge: Descriptive statistics

T1 and T2 data on the CK mathematics assessment was collected for 40 out of the 43 children who consented to participate in individual assessments; 3 children did not complete the full assessment at either T1 or T2.

Table 4.3: Descriptive statistics for Curriculum Knowledge scores for experimental and control conditions at T1 and T2

Condition	Sample	T1 Mean Ability Score (SD)	T2 Mean Ability Score (SD)	Mean Learning Gains (SD)
Control	Nursery A (n = 13)	13.38 (10.99)	17.54 (11.97)	4.15 (5.76)
	Nursery B (n = 8)	8.88 (6.66)	11.75 (8.94)	2.88 (7.20)
	Overall (n = 21)	11.67 (9.65)	15.33 (11.06)	3.67 (6.20)
Experimental	Nursery A (n = 12)	11.75 (8.73)	20.75 (12.68)	9.00 (7.53)
	Nursery B (n = 7)	11.14 (7.08)	17.34 (8.72)	6.29 (3.99)
	Overall (n = 19)	11.53 (7.96)	19.53 (11.24)	8.00 (6.46)

As shown in Table 4.3 and Figure 4.2, mean scores increased for children in both conditions between T1 and T2, but the overall increase in mean score was greater for children in the experimental group. A similar pattern was seen across both nurseries, taking into consideration some differences between conditions at T1 in each setting.

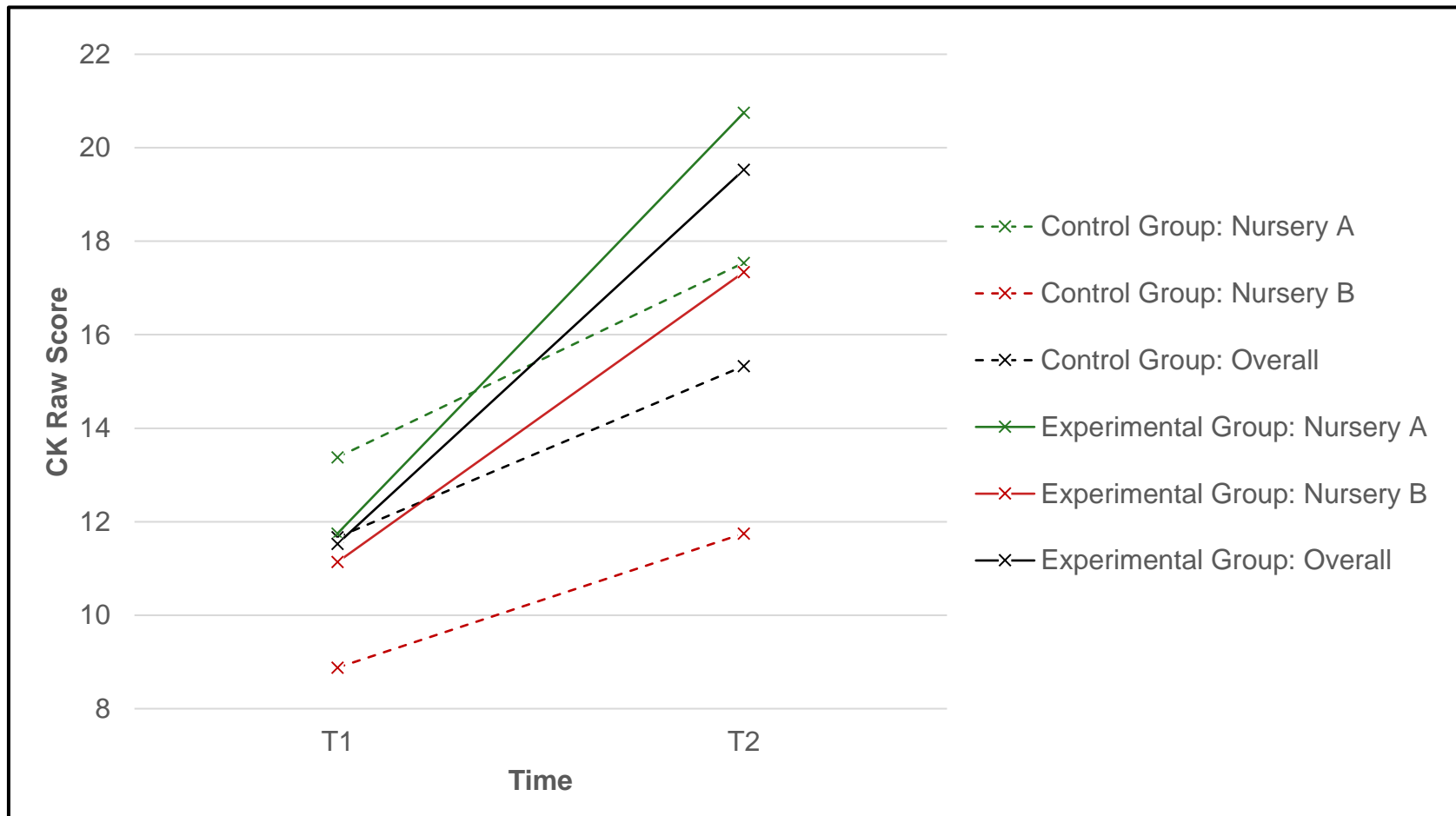


Figure 4.2: A line graph to show change in experimental and control conditions at T1 and T2 in Curriculum Knowledge raw scores

4.1.2.5 Curriculum Knowledge: Inferential statistics

Following preliminary checks (see Appendix 8.16), an ANCOVA was conducted to determine whether there was a statistically significant difference between mathematics CK scores at post-test (DV) between experimental conditions, after controlling for pre-test scores (covariate). The results indicated a statistically significant difference between the experimental and control groups at post-test, after adjusting for pre-intervention scores, $F(1, 37) = 4.58$, $p = .039$, $\eta^2 = .110$. Children in the experimental condition (adjusted $M = 19.60$) scored more highly than those in the control condition (adjusted $M = 15.26$).

A two-tailed independent samples t-test indicated there was not a statistically significant difference between groups on T1 mathematics CK scores, $t(39) = .17$, $p = .864$, again, suggesting that groups were matched for initial mathematical ability on this measure.

4.1.2.6 Curriculum Knowledge: Effect size

An ES analysis, displayed in Table 4.4, indicated that the magnitude of the difference between conditions at post-test was small-moderate.

Table 4.4: Effect size analysis of the difference between conditions in T2 Curriculum Knowledge scores

	Unadjusted means (SD)	Cohen's d	Size of Effect
Total Sample ($n = 40$)	Control: 15.33 (11.06) Experimental: 19.53 (11.24)	+ 0.38	Small-moderate

4.1.3 Data analysis for Research Question 2

A longitudinal follow-up was included in the research design to address Research Question 2:

Are mathematical attainment gains sustained 5 months after the end of the intervention?

The DVs were children's mathematical attainment, as measured by children's ability scores on the mathematics CK assessment at T3¹⁹. It was hypothesised that at T3 children in the intervention group would have significantly higher scores on these assessments, compared to children in the control group.

4.1.3.1 T3 Curriculum Knowledge: Descriptive statistics

At the point of longitudinal follow-up, three children had moved to other settings, leaving 40 children in total. Three children did not complete the full assessment at T1 or T2, and one child was absent at T3, resulting in a final sample size of $n = 36$. Descriptive statistics for the mathematics CK data for this reduced sample at T1, T2 and T3 are displayed in Table 4.5 and Figure 4.3.

¹⁹ Statistically analyses indicated that group differences in ability scores on the ENC sub-test remained non-significant at T3 and therefore they have not been reported here.

Table 4.5: Descriptive statistics for the mathematics Curriculum Knowledge scores for experimental and control conditions at T1, T2 and T3 for the reduced sample

Condition	Sample	T1 Mean Ability Score (SD)	T2 Mean Ability Score (SD)	T3 Mean Ability Score (SD)
Control	Nursery A (<i>n</i> = 11)	13.00 (11.04)	17.55 (12.79)	22.73 (13.29)
	Nursery B (<i>n</i> = 8)	8.88 (6.66)	11.75 (8.94)	18.00 (11.14)
	Overall (<i>n</i> = 19)	11.26 (9.45)	15.11 (11.43)	20.74 (12.33)
Experimental	Nursery A (<i>n</i> = 12)	11.75 (8.73)	20.75 (12.79)	23.75 (13.53)
	Nursery B (<i>n</i> = 5)	11.20 (8.56)	16.80 (9.27)	22.40 (12.70)
	Overall (<i>n</i> = 17)	11.59 (8.41)	19.59 (11.66)	23.35 (12.91)

The descriptive statistics indicate that from T2 to T3, the magnitude of the difference between conditions has reduced. Figure 4.3 indicates that this effect appears to be driven predominantly by a slower rate of learning in the intervention group after T2 at Nursery A. Interestingly, the effect of the intervention appears to be broadly maintained at Nursery B over time, although differences by setting should be interpreted with caution given that the sample size of the intervention group at Nursery B by T3 had reduced to *n* = 5.

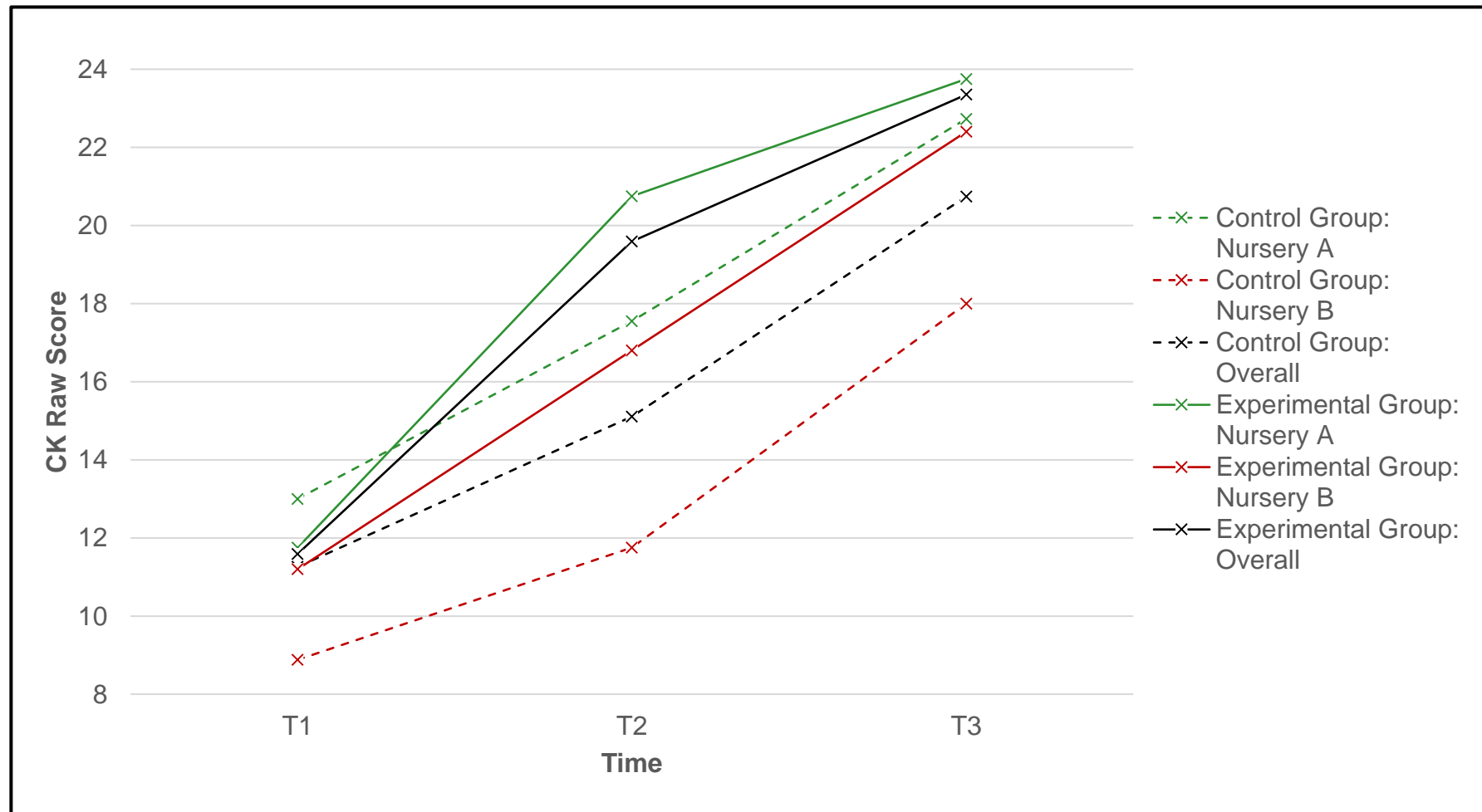


Figure 4.3: A line graph to show change in experimental and control conditions at T1, T2 and T3 in Curriculum Knowledge raw score

4.1.3.2 T3 Curriculum Knowledge: Inferential statistics

Following preliminary checks (see Appendix 8.16), an ANCOVA was conducted to determine whether there was a statistically significant difference between mathematics CK scores at longitudinal follow-up between experimental conditions (DV), after controlling for pre-test scores (covariate). The results indicated that there was not a statistically significant difference between the experimental and control groups at follow-up on the mathematics CK assessment, after adjusting for pre-intervention scores, $F(1, 33) = .68$, $p = .417$, $\eta^2 = .020$. Therefore, it was not possible to reject the null hypothesis for this research question.

A two-tailed independent samples t-test indicated that there was no statistically significant difference between groups on T1 ENC ability scores in this reduced sample, $t(34) = .11$, $p = .914$, indicating that matching at pre-test was not affected by attrition.

4.1.3.3 T3 Curriculum Knowledge: Effect size

An ES analysis, see Table 4.6, indicated that the magnitude of the difference between groups at T2 ($d = + 0.38$) reduced by T3 ($d = + 0.21$).

Table 4.6: Effect size analysis of the difference between conditions in T3 mathematics Curriculum Knowledge

	Unadjusted means (SD)	Cohen's <i>d</i>	Size of Effect
Total Sample (<i>n</i> = 36)	Control: 20.74 (12.33) Intervention: 23.35 (12.91)	+ 0.21	Small

4.1.4 Data analysis for Research Question 3

The third research question that the present study aimed to address was:

What is the impact of the onebillion tablet intervention on the receptive language skills of preschool children aged 3-4 years?

The DV was children's receptive language skills, as measured by the raw total score on the receptive language sub-tests of the CELF-P2 (Wiig et al., 2006). It was hypothesised that at post-test (T2) children in the intervention group would have significantly higher receptive language skills than children in the control group.

4.1.4.1 Receptive language: Descriptive statistics

Out of the 43 participants who consented to participate in individual assessments, one child did not have sufficient knowledge of English animal names at T1 to complete the receptive language test and therefore the assessment was not administered. Another child did not fully complete the assessment at T2. Descriptive statistics for the receptive language data gathered for the remaining 41 participants are displayed in Table 4.7, indicating minimal difference between mean receptive language scores between conditions at T1 or T2.

Table 4.7: Descriptive statistics for receptive language for experimental and control conditions at T1 and T2

Condition	Sample	Pre-test Mean Ability Score (SD)	Post-test Mean Ability Score (SD)	Mean Learning Gains (SD)
Control	Nursery A (<i>n</i> = 13)	34.00 (12.92)	41.15 (10.78)	7.16 (7.29)
	Nursery B (<i>n</i> = 8)	26.00 (13.26)	32.38 (12.15)	6.38 (3.50)
	Overall (<i>n</i> = 21)	30.95 (13.33)	37.81 (11.85)	6.86 (6.03)
Experimental	Nursery A (<i>n</i> = 13)	33.38 (13.46)	39.38 (12.58)	6.00 (5.02)
	Nursery B (<i>n</i> = 7)	27.71 (7.80)	34.86 (13.16)	7.14 (7.49)
	Overall (<i>n</i> = 20)	31.40 (22.89)	37.80 (12.63)	6.40 (5.83)

4.1.4.2 Receptive language: Inferential statistics

Following preliminary checks (see Appendix 8.16), an ANCOVA was conducted to determine whether there was a statistically significant effect of the intervention on receptive language raw scores at post-test (DV) between experimental conditions, after controlling for pre-test scores (covariate). The results indicated that there was not a statistically significant difference between conditions, after adjusting for pre-intervention scores, $F(1, 38) = .05$, $p = .827$, $\eta^2 = .001$. Therefore it was not possible to reject the null hypothesis for this research question.

A two-tailed independent samples t-test indicated that there was no statistically significant difference between groups on T1 receptive language raw scores, $t(40) = .15$, $p = .883$, suggesting that equivalence could be assumed.

An ES analysis was not completed as there was neither a statistically significant effect nor a trend towards an effect.

4.1.5 Data analysis for Research Question 4

The study also aimed to determine:

What is the impact of the onebillion tablet intervention on the AtL of preschool children aged 3-4 years?

Children's AtL was measured by teacher ratings on the Initiative sub-scale of the DECA-P2 (LeBuffe & Naglieri, 2012). It was hypothesised that at T2 children in the intervention group would have significantly higher initiative skills than children in the control group.

4.1.5.1 Approaches to learning: Descriptive statistics

A full set of data was obtained on the AtL 'Initiative' measure for all 47 participating children. Descriptive statistics for this data are displayed in Table 4.8. Preliminary checks indicated that the data were not normally distributed (see Appendix 8.16) and therefore median scores and interquartile ranges (IQRs) have been reported.

Table 4.8 indicates that there were minimal differences between overall median AtL scores at T1 and T2, although there were improvements in both conditions over time. Improvements in AtL over time appear to be driven by changes perceived by the teacher at Nursery A; scores remained relatively stable at Nursery B.

Table 4.8: Descriptive statistics for the approaches to learning for experimental and control conditions at T1 and T2

Condition	Sample	Pre-test Median Score (IQR)	Post-test Median Score (IQR)
Control	Nursery A (<i>n</i> = 14)	21.0 (9)	27.0 (6)
	Nursery B (<i>n</i> = 10)	23.0 (18)	23.5 (17)
	Overall (<i>n</i> = 24)	22.5 (12)	27.0 (12)
Experimental	Nursery A (<i>n</i> = 13)	21.0 (11)	27.0 (10)
	Nursery B (<i>n</i> = 10)	23.5 (19)	23.5 (21)
	Overall (<i>n</i> = 23)	21.0 (13)	27.0 (10)

4.1.5.2 Approaches to learning: Inferential statistics

As the data were not normally distributed, a combination of non-parametric tests were conducted, using Wilcoxon Signed-Rank tests, for within-subjects contrasts, and Mann Whitney U-tests, for between-subjects contrasts.

Within-subjects contrasts: Data for each condition were compared between T1 and T2 for each condition using a two-tailed Wilcoxon Signed-Rank test. The purpose was to identify whether there were any significant changes in children's Initiative scores over time. In the control condition, there was a statistically significant improvement in children's scores between T1 and T2, $Z = 2.99$, $p = .003$. In the experimental condition, there was also a statistically significant increase in children's scores over time, $Z = 2.12$, $p = .034$.

Between-subjects contrasts: Two-tailed Mann-Whitney U tests revealed no significant difference between experimental and control conditions at T1 ($Z = .51$, $p = .609$) or T2 ($Z = .33$, $p = .741$). Therefore the null hypothesis for Research Question 4 could not be rejected.

Again, an ES analysis was not conducted as there was no significant intervention effect.

4.1.6 Summary of quasi-experimental findings

In summary, there were mixed findings in this study concerning the impact of the *onebillion* intervention on mathematical attainment, with statistically significant intervention effects identified on a researcher-developed measure of mathematics CK (Outhwaite et al., 2017), but no statistically significant impact on a standardised measure of numerical conceptual understanding, the ENC sub-test (Elliot & Smith, 2011). ES analysis indicated a larger effect on the CK assessment ($d = +0.38$), than on the ENC sub-test ($d = +0.25$). The intervention did not, however, have a statistically significant impact on mathematics attainment at 5 month follow-up, although mean CK assessment scores were higher for the experimental group compared to the control group (between-group ES: $d = +0.21$). There was no statistically significant intervention effect on broader areas of development, including receptive language and AtL at post-test.

4.2 Embedded Data Analysis

This section of the results focuses upon analyses of data from the embedded aspects of the design, addressing Research Questions 5 and 6:

5.) What are facilitators' perceptions of the outcomes of the onebillion intervention for preschool children aged 3-4 years?

6.) What factors may affect the outcomes of the onebillion intervention for preschool children aged 3-4 years?

The qualitative data analyses of the semi-structured interviews and narrative observations are presented first. Additional quantitative analyses of the structured observation data and correlations between measures are then presented to further explore Research Question 6.

4.2.1 Qualitative data analysis procedures

4.2.1.1 Thematic analysis

Thematic analysis (TA) was selected as the most appropriate method for analysing qualitative interview and narrative observational data, given that it is a highly flexible method of identifying themes (patterns of meaning) across a data set in relation to a particular research question (Braun & Clarke, 2006; Clarke & Braun, 2017). Unlike other methods of qualitative analysis, TA can be applied to a wide range of different types of data and makes no implicit assumptions about the researcher's particular theoretical or philosophical positioning (Braun & Clarke, 2013). TA is therefore compatible with the pragmatic stance adopted by the researcher and applicable to both types of qualitative data collected during the study.

TA can be conducted using an inductive 'bottom up' approach, exploring themes closely linked to the data-set itself, or a more 'top down' approach, where the researcher is guided by particular theoretical content (Braun & Clarke, 2006). In this case, the researcher aimed to illuminate participants' own experiences and perceptions through a 'bottom-up' data-driven approach to the development of themes. Nevertheless, it is acknowledged that the researcher's identification of themes is likely to some extent have been influenced by their prior theoretical knowledge and experience in this area.

Themes can also be identified at either a 'semantic' level, focusing on identifying themes within the explicit meaning of the data, or a 'latent' level, focused on identifying patterns within the underlying ideas and assumptions of the data (Braun & Clarke, 2006). In this study, the researcher identified themes at a 'semantic' level before moving to greater interpretation at a later stage in the discussion of findings, when themes were related to previous literature and theory.

The process of TA was guided by the six step process outlined by Braun and Clarke (2006), as follows:

1. *Data Familiarisation*: All facilitator interviews were transcribed and narrative observations typed from field notes. The researcher read through the entire data set, noting down any salient features.
2. *Initial Coding*: The researcher coded all extracts from the data set that were relevant in answering research questions 5 and/or 6.
3. *Searching for Themes*: Initial codes were then collated into potential themes.
4. *Reviewing Themes*: Themes were reviewed to ensure that they worked within the coded extracts and the remaining data set as a whole.
5. *Defining and naming themes*: Clear definitions and names were given to capture the essence of each theme.
6. *Producing the report*: An analytic narrative was written, capturing the story of the data and answering the research question(s).

TA has been critiqued for losing the individual 'voice' of each participant by focusing too greatly on searching for themes (Braun & Clarke, 2013). For this reason, contradictions across participants and between observation sessions have been identified and highlighted within the analytic narrative in Sections 4.2.3-4.2.4. TA was conducted separately for both the interview and observation data, but the findings were compared and contrasted in stages 5-6 to support triangulation.

4.2.1.2 Ensuring quality and trustworthiness

Paralleling standards of reliability and validity in quantitative research, a number of separate criteria have been established for judging the quality and trustworthiness of qualitative research (Guba & Lincoln, 1989). Steps taken in addressing each of these criteria within qualitative aspects of the research design are displayed in Table 4.9.

Table 4.9: Quality and trustworthiness of qualitative aspects of the design

Criteria Definition	Steps Taken to Address Quality and Trustworthiness
<p>Credibility</p> <p><i>Similar to internal validity; reflecting the extent to which the hypotheses developed are a reflection of the phenomenon or experiences of interest and can be supported by the data set.</i></p>	<ul style="list-style-type: none"> • In order to ensure sufficient supporting data for any claims, 12 observations of intervention sessions were conducted (including at least one observation with each facilitator/group of children), spaced throughout the intervention period. All facilitators were interviewed. • Facilitators confirmed that interview transcripts were accurate and member checks of identified themes/subthemes were conducted with 3 out of 4 facilitators. • The researcher received regular supervision during design/analysis phases. • The researcher monitored her own developing constructions throughout the study, maintaining a reflexive research journal. The researcher's positionality (beliefs, values and experiences) in relation to the research context were discussed in supervision. The researcher's prior teaching experience and work as a TEP will have influenced the data gathered and the interpretations drawn. • Triangulation was possible through comparison of TA from interviews/observations.
<p>Transferability</p> <p><i>Similar to external validity; reflects the extent to which readers of the research can identify the applicability of the research to other contexts</i></p>	<ul style="list-style-type: none"> • Detailed thick description of the children, facilitators and schools participating in the research has been given, as well as implementation procedures. • The research was conducted over two different settings and involved four different groups of children/facilitators. The research was, however, conducted on a small scale, affecting the applicability of findings to other contexts.

Criteria Definition	Steps Taken to Address Quality and Trustworthiness
<p>Dependability</p> <p><i>Similar to reliability; reflecting the consistency in use of different processes in the research, supporting replication</i></p>	<ul style="list-style-type: none"> • All processes in data collection are clearly stated. • Inter-rater reliability checks with two other TEPs were conducted (see Section 4.2.2.4).
<p>Confirmability</p> <p><i>Similar to objectivity; reflecting the extent to which themes and interpretations can be linked to data</i></p>	<ul style="list-style-type: none"> • A clear chain of evidence is presented in the stages of TA. • Interpretations are linked clearly to data extracts. • Inter-rater reliability checks were conducted.
<p>Authenticity</p> <p><i>Reflects the extent to which all views are represented fairly</i></p>	<ul style="list-style-type: none"> • Contradictions in the accounts are explicitly stated. • Data extracts explicitly linked to each theme to support a trail between each participant's voice and interpretations reached. • Member checks of TA from interview transcripts to allow participants time to comment on the identified themes.

4.2.2 Data analysis for Research Questions 5 and 6

4.2.2.1 Step 1: Data familiarisation

Data familiarisation involved transcribing audio recordings from the interviews (see notation system in Appendix 8.17) with the facilitators and typing up written notes from the narrative observations (see Table 4.10 for further details of observed sessions). The researcher then re-read all data transcripts/observations, identifying any salient points which might inform later stages of analysis.

Table 4.10: Narrative observations conducted during intervention sessions

Observation Identifier	Group	Facilitator	Number of Children Present
A and B	<i>Nursery B – Morning Session (two groups)</i>	Danielle	3 and 4
C	<i>Nursery A – Afternoon Session</i>	Bianca	3
D	<i>Nursery A – Afternoon Session</i>	Anne	4
E and F	<i>Nursery B – Morning Session (two groups)</i>	Clara	5 and 4
G	<i>Nursery A – Morning Session</i>	Anne	8

4.2.2.2 Step 2: Initial coding

The purpose of initial coding was to identify all data extracts across both interview transcripts and narrative observations that would be relevant in answering Research Questions 5 and 6. Codes are a “short word or phrase that symbolically assigns a summative, salient, essence-capturing and/or evocative attribute for a portion of language based...data” (Saldaña, 2016, p.4) Consistent with a data-driven approach, a complete coding procedure was used (Braun & Clarke, 2013), whereby the researcher systematically works through the full data set, coding any portions of data that are relevant in answering the researching questions rather than selectively coding pre-determined phenomenon of interest. Multiple codes were assigned to a particular data extract if appropriate to ensure that all potentially relevant aspects of the data

were captured. An example of data coding for a section of an interview transcript is given in Appendix 8.18.

In order to support later identification of themes, initial codes were then reviewed to reassign overlapping codes and to remove any codes that were unique to only one facilitator or observation, unless they directly contradicted others. This process resulted in 58 final codes for the interview transcripts (Appendix 8.19) and 56 final codes for the narrative observations (Appendix 8.20). Data extracts assigned to each code were then collated to support step 3 of the TA.

4.2.2.3 Step 3: Searching for themes

Following coding, the researcher progressed to identifying themes within each data set. Recorded codes were organised into 'candidate themes' (Braun & Clarke, 2013), which were groups of codes that appeared to cluster together around a similar idea. As recommended by Braun and Clarke (2013), codes that did not appear to fit any candidate themes were retained at this stage under a category of 'Miscellaneous' should they become relevant when themes were refined.

Interview data extracts were collated into themes which addressed Research Question 5 (*What are facilitators' perceptions of the outcomes of the onebillion intervention for preschool children aged 3-4 years?*) and Research Question 6 separately (*What factors may affect the outcomes of the onebillion intervention for preschool children aged 3-4 years?*). Narrative observation data extracts were, however, only relevant in answering Research Question 6. The initial themes and subthemes identified at this stage from each TA are displayed in Appendices 8.21 and 8.22.

4.2.2.4 Step 4: Reviewing themes

The purpose of reviewing the candidate themes at this stage was to check consistency with the coded data extracts and the rest of the data set (Braun & Clarke, 2013). First, the researcher reviewed all coded data to ensure that each theme was coherent and distinct from other themes (Braun & Clarke, 2006). Second, candidate themes were checked with the remainder of the data set to ensure that they captured its overall meaning and to recode any missed data

extracts (Braun & Clarke, 2006). Following this process, the researcher discussed the final identified themes, below, in supervision and inter-rater reliability checks and member checks were conducted (see Table 4.9).

Inter-rater reliability checks were conducted following a method outlined by Joffe (2012). Two other TEPs allocated 10% of relevant data extracts to one of the final themes and/or subthemes that were identified (see example in Appendix 8.23). For the narrative observations, there was 78% agreement between the researcher and TEP 1 and 82% agreement between the researcher and TEP 2. For the interviews, there was 75% agreement between the researcher and TEP 1 and 79% agreement between the researcher and TEP 2. In each case, concordance was greater than or equal to 75%, indicating a dependable analysis.

During member checks, each facilitator reported that the overall themes/subthemes identified from analysis of the interview transcripts reflected their views; some additional comments were also made at this stage, including contradictions to particular subthemes, and these were used to inform the analytic narrative in Sections 4.2.3 and 4.2.4.

Although the process of TA was completed separately for the narrative observations and interviews, there was overlap in all of the main themes and some of the subthemes identified in the data sets for Research Question 6. Themes are therefore presented and discussed collectively in the remaining sections. Overlap between the main themes identified provides triangulation about potential factors that may have influenced the outcomes of the intervention.

4.2.2.5 Step 5: Defining and naming themes

The last step of TA was to fully define each of the identified themes, capturing the “essence of what each theme is about” (Braun & Clarke, 2006, p. 92). The following sections capture the story told by each theme and the data as a whole in relation to Research Questions 5 and 6. Pseudonyms for facilitators and children are used throughout this account to protect anonymity.

4.2.3 Research Question 5: Key themes

Research question 5 asked, “*What were facilitators’ perceptions of the outcomes of the onebillion intervention for preschool children aged 3-4 years?*”

Four broad themes were identified from the facilitator interviews:

- Individual differences
- Improvements in mathematical knowledge and skills
- Developed confidence
- Developed attention skills

A thematic map displaying these themes and related subthemes is shown in Figure 4.4. Each theme is further defined and exemplified in this section.

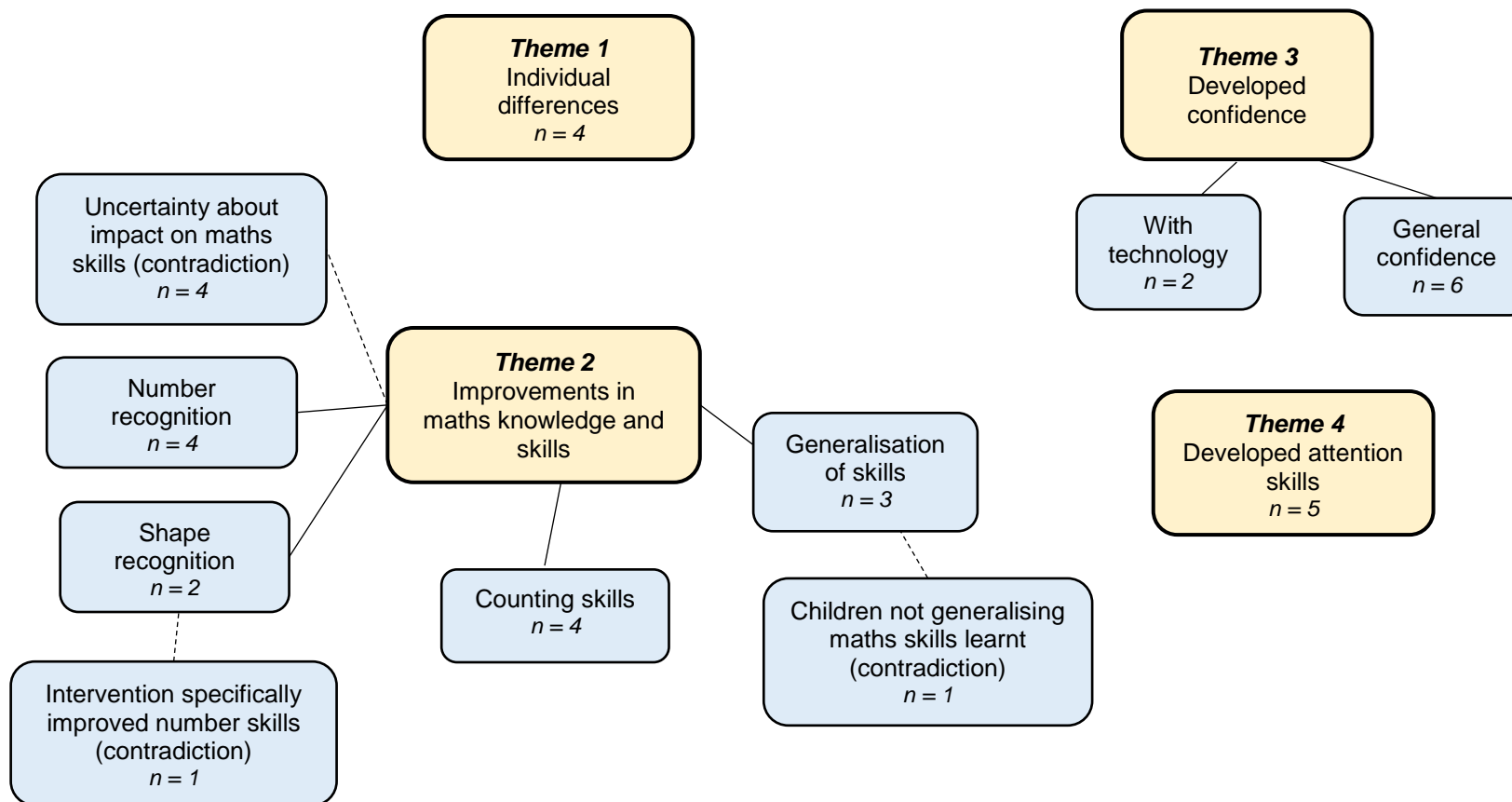


Figure 4.4: Thematic map for Research Question 5: What are facilitators' perceptions of the outcomes of the intervention? The number of data extracts relevant to each theme/subtheme is given in italics.

4.2.3.1 Theme 1: Individual differences

Most facilitators felt that the outcomes of the intervention had not been uniform across the group and were, *“different for the different children.”*^(Clara) The perception of individual differences in Rtl affects the interpretation of the group level quantitative analyses presented previously, suggesting that what may ‘work’ for one child may not work for all. Potential factors affecting Rtl are explored in Section 4.2.4.

4.2.3.2 Theme 2: Improvements in mathematical knowledge and skills

In general, facilitators perceived the intervention to have positive outcomes for the mathematical knowledge and skills of some children, particularly in the area of number sense, although views were more divergent about the impact on shape. Anne stated, *“[The children] seem to have been getting further now with their counting [...] and starting to recognise a few more numbers.”* Children were also thought to be applying newly acquired skills in other learning activities: *“[...] they’ve come away after using the app and used whatever they’ve done on the app in nursery.”*^(Anne) Use of the intervention therefore was perceived to support generalisation of skills to new contexts.

However, some contradictions were apparent. For example, Clara suggested that learning gains may have simply been *“a natural improvement”* over time, due to maturation effects, and felt that children were not *“necessarily generalising [skills] off the apps.”* Variation in perception might reflect differences in children’s Rtl, as discussed above, and the type of skill learnt, see Chapter 5.

4.2.3.3 Theme 3: Developed confidence

All facilitators identified improvements in some children’s confidence in using technology, as well as greater overall self-confidence for some individuals. Clara commented:

“Matthew [...] benefited in a different way because he was quite low in confidence [...] but [...] because he was making progress through it, he was really enjoying that [...] It was almost like it was giving him a little boost [...] He’s just more confident in the [...] classroom and in the environment.”

This theme indicates that the intervention was perceived to have benefited some children's confidence in approaching other activities in the nursery – one aspect of AtL also measured by the teacher-report 'Initiative' scale used in this study (LeBuffe & Naglieri, 2012).

4.2.3.4 Theme 4: Developed attention skills

It was also felt that the intervention improved some children's attention. Anne felt that the intervention helped to "*build up on that concentration time and actually focus at an activity.*" Task engagement is another important aspect of AtL (McClelland, Acock, & Morrisson, 2006), and features of the intervention that may have affected this outcome are explored further in Section 4.2.4.

4.2.3.5 Research Question 5: Summary

Overall, some children were perceived to have benefited from the intervention more than others. Nonetheless, facilitators generally felt that the intervention had benefited at least some children's mathematical skills and knowledge, particularly within the domain of number. Some facilitators also perceived additional benefits for some children's confidence and attention in learning environments.

4.2.4 Research Question 6: Key themes

Data extracts from both facilitator interviews and narrative observations were used to address Research Question 6, “*What factors may have affected the outcomes of the onebillion intervention for preschool children aged 3-4 years?*”

Five broad themes were identified:

- Language demands of the intervention
- Children’s attention on-task
- Pedagogy and instructional level of the intervention
- Children’s attitude to learning
- Implementation in the nursery setting

Each of these themes are defined and exemplified in this section, alongside thematic maps depicting each of the identified subthemes. Closely related or directly contradictory subthemes shown in the diagrams are discussed collectively under key subheadings in the following sections.

4.2.4.1 Theme 1: Language demands of the intervention

First, a key set of factors influencing the outcomes of the intervention were related to the language demands of the *onebillion* software. A map of key subthemes related to this theme is displayed in Figure 4.5.

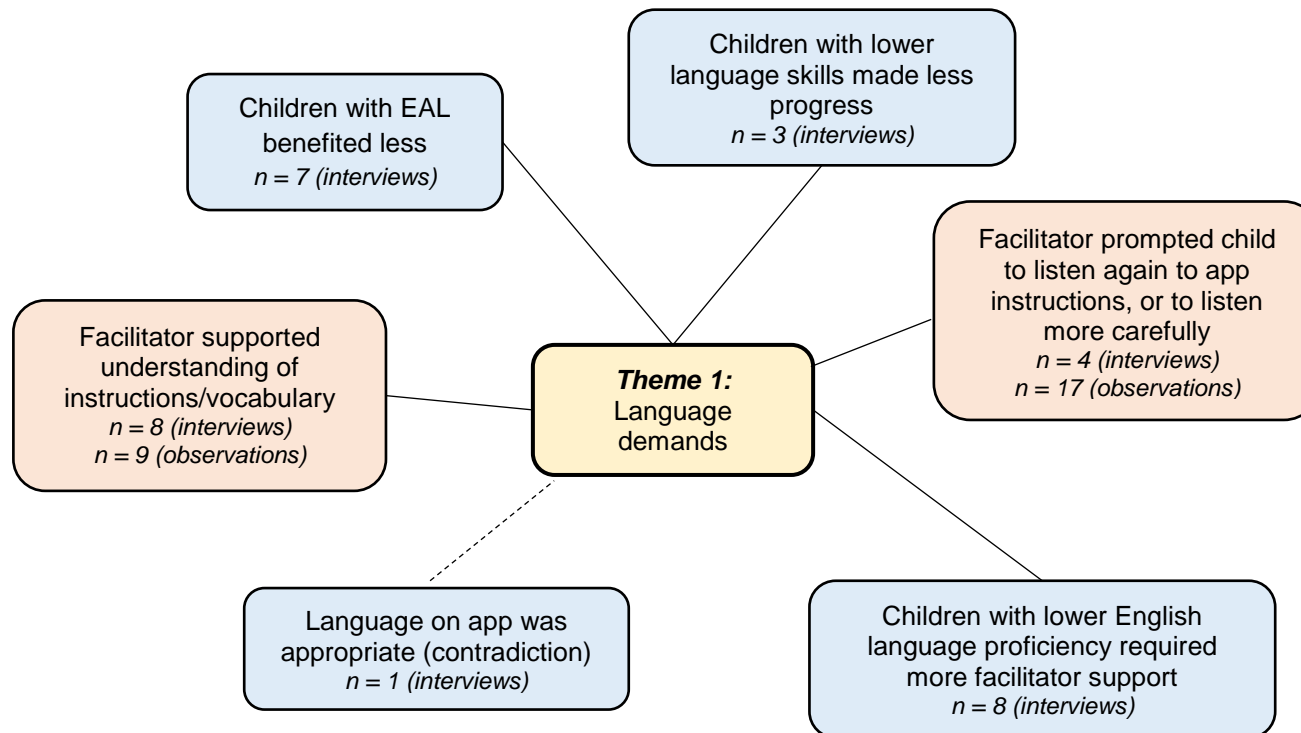


Figure 4.5: Thematic map for Theme 1: Language demands of the intervention subthemes

The number of data extracts relevant to each subtheme is given in italics.

Red = identified from both narrative observations and interviews

Blue = identified from interviews only

Language proficiency:

There was a perception that children with lower language skills, including those learning EAL, did not benefit as much from accessing the intervention. Clara commented, “[*Our EAL children*] found it difficult to follow the instructions and even though it was as visual as it can be, [...] I think [...] it is still quite language based.” The app software relies upon children listening and understanding instructions provided by the virtual teacher, which, for some, may have been a barrier to learning.

Facilitator role:

These children seemed to be provided with higher levels of facilitator scaffolding: “[*The EAL children*] were the ones who tended to have most [...] support.”^(Clara) Facilitator support appeared to involve asking children to use software features to repeat instructions from the virtual teacher, as well as rephrasing instructions: “I was [...] getting them to just listen again and when possible making the vocab easier.”^(Anne) This type of facilitator support was also frequently seen by the researcher during observations of intervention sessions.^(e.g. Observation E)

Contradiction:

Clara, however, did comment, “I think the language that [...] was used [*in the apps*] was actually quite appropriate [...] for the children,” indicating that in general the language may have been aligned with the broader curriculum and attuned to the needs of many children, but perhaps not consistently and therefore potentially leading to individual differences in outcomes.

4.2.4.2 Theme 2: Children’s attention on-task

An additional factor that may have influenced the success of the *onebillion* intervention was the extent to which children sustained attention on-task during the intervention. A thematic map of subthemes for *Theme 2* is displayed in Figure 4.6.

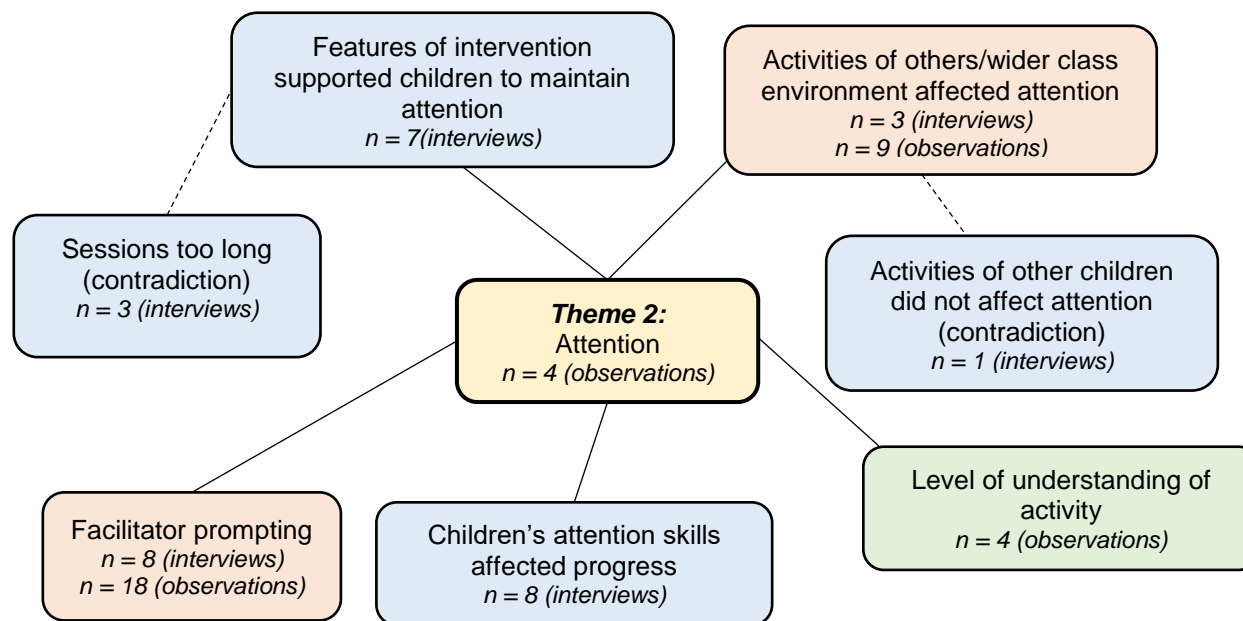


Figure 4.6: Thematic map for Theme 2: Children's attention on-task subthemes

The number of data extracts relevant to each subtheme or theme is given in italics.

Red = identified from both narrative observations and interviews

Blue = identified from interviews only

Green = identified from narrative observations only

Attention skills:

Data suggested that some children's attention skills were felt to have influenced intervention gains: "*Cameron [has] struggled a lot. [...] his attention is very poor.*" (Bianca) Where children found it difficult to maintain attention, it is possible that they made less progress through the intervention; children's level of progress through the app activities was related to learning gains, as discussed below (see Section 4.2.5.2).

Class environment:

Perceptions of the impact of the class environment on children's attention on-task differed across settings. At Nursery B, Danielle noted, "*[Other children in the class] would keep coming to you, which was a distraction because [...] all of a sudden they'd look up [...] to see who's come to [...] ask questions.*" During observations at Nursery B, the researcher also noted environmental influences on children's concentration, including loud noises. (e.g. Observation B) These factors, however, were perceived to have had less impact during the smaller afternoon sessions at Nursery A: "*I mean the noise levels and things like that, other activities going on, didn't bother them.*" (Bianca) Variation in nursery context, perhaps including staffing levels and class size, may have influenced outcomes.

Intervention features:

There were divergent views about how features of the intervention supported attention. Bianca commented, "*[Activities] were quite varied and I think that held their concentration a little while longer,*" whilst some facilitators felt that the intervention sessions could have been shortened: "*I'm not sure whether fifteen minutes is a bit long for the nursery children.*" (Anne) Some aspects of the intervention, particularly software content, may therefore have enhanced learning gains, but fifteen minutes may have placed too many demands on attention.

Facilitator role:

All facilitators felt that they had a role in supporting children's concentration during intervention sessions. For example, Bianca stated that she "*was trying to keep them on-task,*" and, on many occasions, facilitators were directly observed prompting children to concentrate. (e.g. Observation C)

Level of understanding:

At times, however, children may have simply required greater instructional support. For example, children were sometimes observed losing focus on-task when they did not appear to understand the activity. (e.g. Observation C) The potential impact of the level of instructional scaffolding that children received is discussed further below.

4.2.4.3 Theme 3: Pedagogy and instructional level of the intervention

The third set of factors that may have influenced the success of the intervention were related to the level of the instruction and pedagogical factors (Figure 4.7).

.

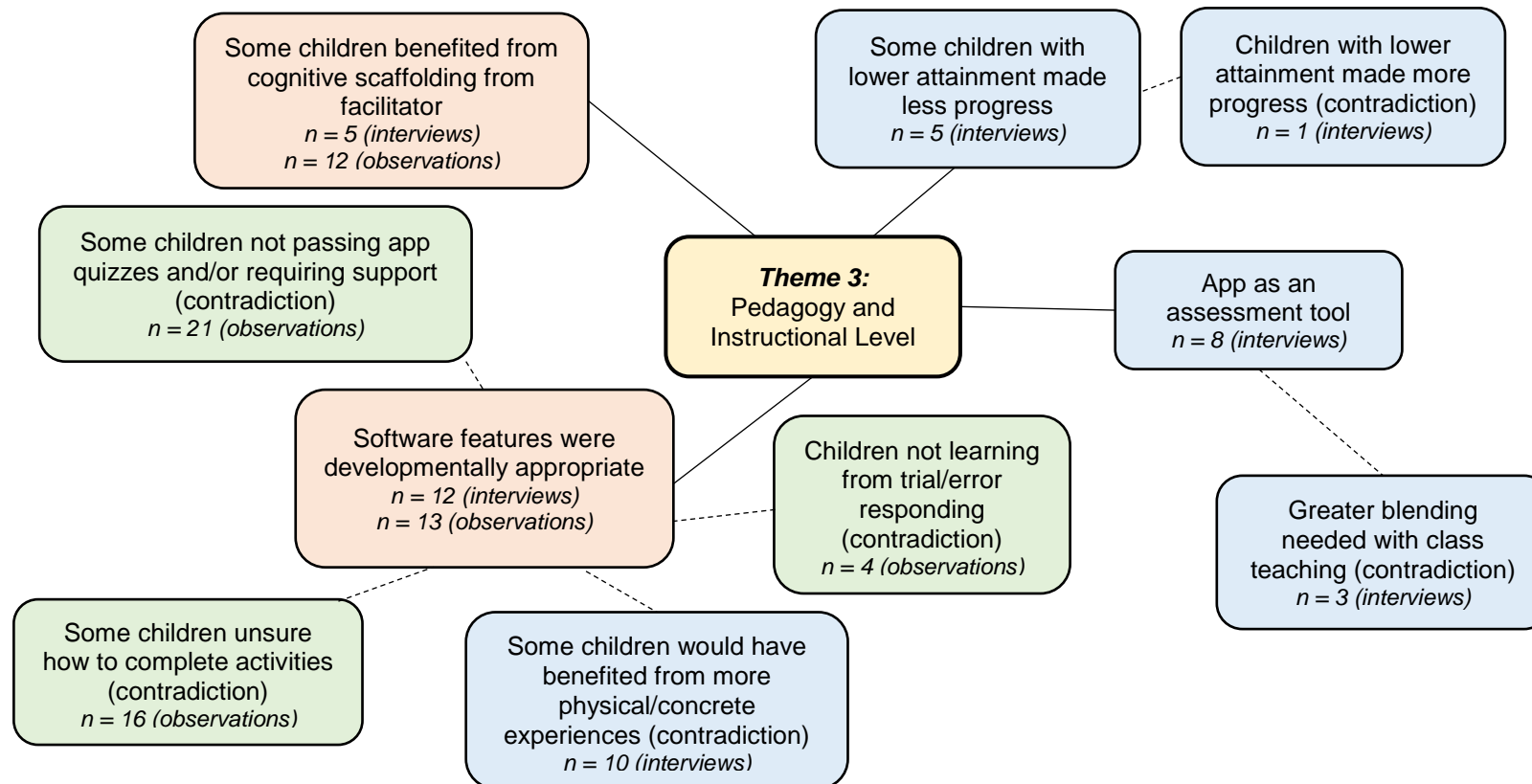


Figure 4.7: Thematic map for Theme 3: Pedagogy and instructional level of the intervention subthemes

The number of data extracts relevant to each subtheme is given in italics.

Red = identified from both narrative observations and interviews

Blue = identified from interviews only

Green = identified from narrative observations only

Prior attainment:

Although there were some divergent comments about how prior mathematical attainment may have influenced outcomes, there was an overall perception that many children with lower initial skills and knowledge made less progress: *“I would say that the more able children have probably had a more positive outcome [...] than the less able children.”*^(Clara) Consequently, it is possible that the instructional content and features of the software may not have sufficiently scaffolded the learning of some children.

Facilitator support:

The level of pedagogical support that children received from facilitators appears to have been an influential factor. All facilitators reported occasions when they needed to provide additional layers of cognitive scaffolding: for example, *“If he’d got to find a number 4 for instance we’d start at the beginning and count through ‘til he got to it.”*^(Bianca) This was also apparent during observed intervention sessions; for example, a facilitator modelled how to draw numerals in the correct direction.^(Observation B) Use of technology in a social context, may therefore have been influential to outcomes, allowing instruction to be more finely-tuned to need.

Instructional level of the software:

Whilst some learners may have needed additional support, the majority of facilitators felt that aspects of software content were developmentally appropriate, including familiar objects and visual features: for example, *“It’s real things for them and everyday things that they’re used to. That was nice.”*^(Anne). Moreover, the researcher often observed children moving through the app activities successfully and independently^(e.g. Observation A), or learning from trial and error approaches^(e.g. Observation D), indicating that the instructional features of the app, such as the staged curriculum, modelling from the virtual teacher and feedback, appeared to be supporting learning at these times.

Nevertheless, the researcher also noted a number of occasions where children appeared unsure how to complete an activity^(e.g. Observation C) and, at times, adopted a meaningless trial and error strategy.^(e.g. Observation G) The researcher also observed that children frequently failed the quiz at the end of the app,

sometimes on multiple occasions, or required additional facilitator support with this aspect of the intervention.^(e.g. Observation F) The software therefore did not always appear to fully secure children's knowledge of mathematical topics and enable them to experience success.

A distinction was also noted between the 2-dimensional experience of the app, and the 3-dimensional physical world. Clara commented, *“[In the typical curriculum] you initially start off with the children counting physical objects that they can move [...] whereas on that app they are asked to count things that are still [...] and that's not necessarily always easy for them to do.”* Some children may therefore have benefited from additional concrete learning experiences.

The relationship between the app and the wider curriculum:

Data also suggested the utility of the app to inform curriculum processes in the classroom. For example, facilitators noted that the apps were an assessment tool to inform other forms of mathematics instruction at home and school:

“If parents asked I [...] said anything they were struggling with so that they could obviously support them at home and it did sort of help me pick up on anything that I needed to probably work on a bit more with the children.” (Anne).

Nevertheless, they felt that it would have been beneficial if the intervention had been more closely blended with the rest of the mathematics curriculum. Danielle stated, *“We could [...] teach them the shape beforehand, then when they went on the iPad to do that activity [...] it would be another way of reinforcing it.”*

4.2.4.4 Theme 4: Children's attitude to learning

A further theme identified from both data sets related to the potential impact of affective factors and children's attitude to learning (see Figure 4.8).

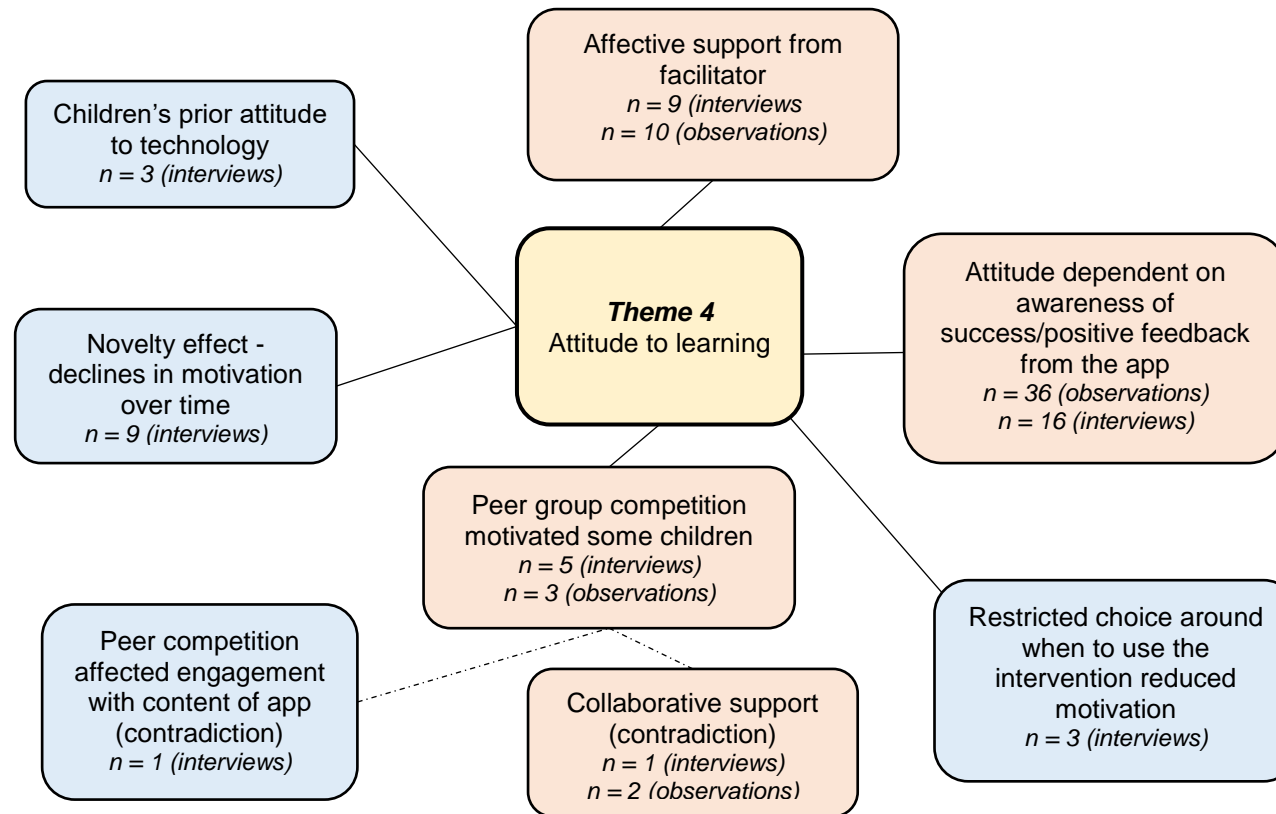


Figure 4.8: Thematic map for Theme 4: Children's attitude to learning subthemes

The number of data extracts relevant to each subtheme is given in italics.

Red = identified from both narrative observations and interviews

Blue = identified from interviews only

Response to feedback and progress:

Children's attitude to learning appeared to be affected by the level and nature of the feedback they received. For example, the researcher observed that positive feedback (e.g. certificates/stars/comments from others) and awareness of progress through the staged curriculum of the app seemed to increase children's enjoyment, confidence and motivation:

Ben cheered when he completed an activity, before remarking, "I just need one more and then I'm on the test." (Observation G)

After receiving a certificate for successfully completing an app quiz, Dean walked around the classroom and showed the iPad to every adult in the room. (Observation A)

Similarly, facilitators also reported that attitude to learning was enhanced from receipt of positive feedback: *"They were so happy and they were cheering, every time they were getting it and going, 'Yesss!'"* (Danielle) Feedback therefore seemed to be enhancing children's feelings of competence.

Some children's awareness of lack of progress, however, was perceived to have a negative impact on attitudes, including motivation: *"Children who weren't necessarily making the progress through [...] were then getting a bit frustrated [...] and disheartened."* (Clara) They also appeared to be showing some hesitancy in responding, perhaps trying to avoid negative feedback and 'failing' app quizzes: for example, one child counted out loud several times to check her answers before responding. (Observation G) Inclusion of negative feedback (an 'error' sound) and the need to achieve a 100% pass rate on app quizzes may therefore have led to less positive attitudes to learning for some children, potentially affecting their Rtl.

Facilitator role:

Facilitators had an important role in providing affective support, including reassurance and encouragement, in order to promote positive attitudes to learning, potentially compensating for some of the effects noted above. Danielle remarked about one child, *"She was just sat there thinking 'Shall I try it?' and then she would try but she wasn't sure if she was right [...] she needed that support."* Some facilitators were also observed directly praising progress, or

encouraging children to persist: for example, “*Never mind, let’s have another go.*” (Anne in Observation A)

Peer group influences:

All of the facilitators identified an element of peer competition between children. Bianca stated, “*It became like a challenge between themselves really, ‘Oh you’re on that one. [...] Well I need to get on to that one as well then’. [...] It made them want to work.*” Children were also directly observed making comparisons between each other about their progress: for example, *after passing an app quiz, David said “I won” and cheered. Another child commented, “You’ve got the same as Luke now.”* (Observation C) Social comparisons therefore appeared to be increasing children’s motivation to learn, perhaps leading to a greater rate of progress through the app and higher learning gains.

Nevertheless, some facilitators felt that that peer competition may have restricted children’s engagement with the instructional content of the app, saying, “*Some of them were so keen on getting through to beat the others that they weren’t necessarily [...] focusing on whether it was right or wrong.*” (Clara)

There was also a perception, particularly at Nursery A, that some peer interactions were more collaborative: “*If there was a child that was next to somebody that didn’t know their numbers they were helping each other.*” (Anne)

The researcher observed children supporting each other with activities (Observation C), and praising each other’s achievements (Observation G) at Nursery A. At times, peer support may therefore have provided additional affective and instructional support for some children encountering difficulties.

Additional factors:

A number of other factors may have influenced children’s motivation using the app. For example, children were thought to be less engaged over time: “*Toward the end [...] some of the children got a bit fed up with it.*” (Anne) An initial novelty effect may therefore have led to greater progress and learning gains in the initial weeks of the intervention.

Some, although not all, facilitators felt that children may have been more motivated to learn if they had more choice about when to use the intervention.

Bianca remarked, *“If they go to choose to use it they don’t realise that they’re learning.”*

Finally, there was a perception that children’s prior attitude to technology may have affected their level of engagement. For example, when discussing a child who benefited less from the intervention, Danielle said, *“I don’t think she really goes on iPads either. [...] She doesn’t seem to be very keen to go and do anything on them.”*

4.2.4.5 Theme 5: Implementation in the nursery setting

A final overall theme that was identified from the data sets were factors relating to implementation of the intervention (see Figure 4.9).

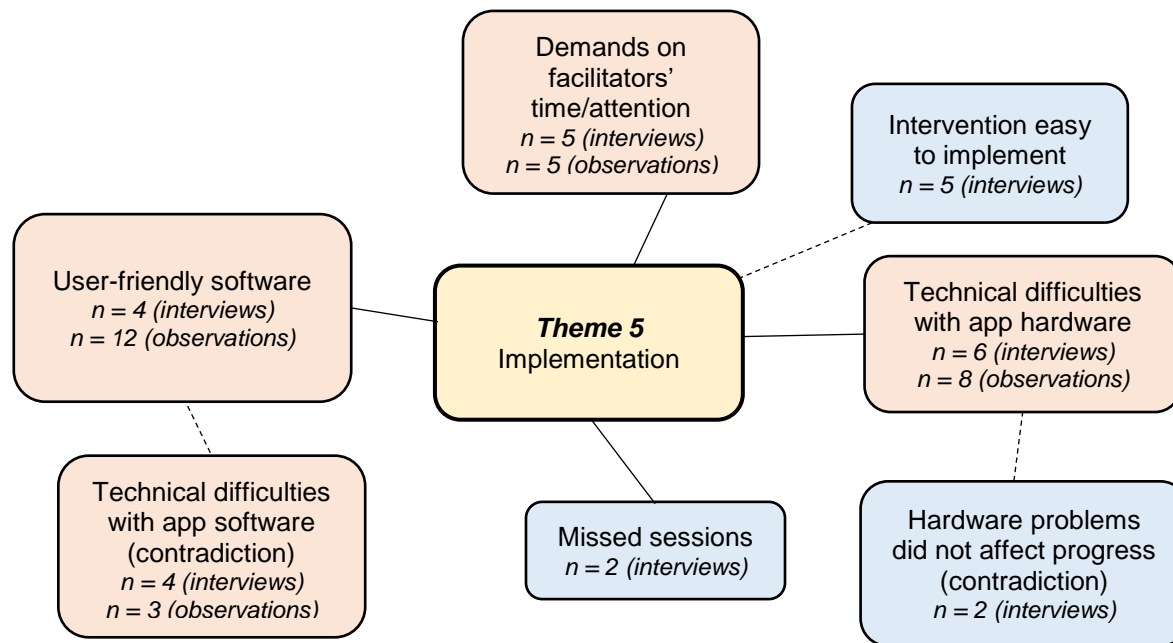


Figure 4.9: Thematic map for Theme 5: Implementation in the nursery setting subthemes

The number of data extracts relevant to each subtheme is given in italics.

Red = identified from both narrative observations and interviews

Blue = identified from interviews only

Software:

Data indicated that the app software was perceived to be user-friendly for the children: *“It was very easy to use, very easy to navigate.”* (Bianca) The researcher also observed many occasions where children were able to navigate independently through the app activities. (e.g. Observation G) This may have reduced the level of technical support that children required, facilitating overall ease of implementation, and allowing children greater autonomy. There was, however, one ‘bug’ within the software that caused the activity to freeze and facilitators had to provide some technical support until this was resolved: *“We had to keep going backwards and trying to sort that out.”* (Bianca)

Hardware:

Technical difficulties with hardware, such as iPad tablets and headphones, appeared to be a more influential factor at Nursery B than Nursery A. For example, Danielle said, *“The batteries of the iPads weren’t charging properly [...] and then sometimes in the middle of a programme one would go off.”* The researcher directly observed children experiencing technical difficulties at that setting. (e.g. Observation B) Where facilitators were provided enhanced technical support they may have had less time to provide other forms of scaffolding, potentially reducing overall learning gains.

Facilitators at Nursery A did not feel that hardware difficulties had impacted on children’s progress. Anne commented, *“A couple of the apps broke with the headphones so we had to use them without headphones but that was no problem. They just [...] sat at opposite sides of the carpet and they could hear well.”* Again, this indicates the potential influence of nursery context on outcomes.

Missed sessions:

Some facilitators noted that children sometimes missed sessions and that the intervention could not run every day as planned, which may have affected some children’s progress. Anne reported, *“We had a few that were away for quite a number of days.”* Clara said, *“We’ve had [...] INSET days and training.”*

Demands on facilitators’ time and attention:

The majority of facilitators felt that demands on their time and attention may

have limited the amount of additional support they could provide children: Danielle noted particular difficulties when staffing was reduced elsewhere in the classroom, saying “*Monday, Tuesday, [other children] would just keep coming every time. [...] You wouldn’t [...] necessarily always be with the group.*” This was observed directly on a number of occasions. (e.g. Observation B) Levels of facilitator instructional support therefore appeared to be influenced by factors such as group size and staffing levels. Danielle also explained that completing intervention sessions impacted on the delivery of other aspects of the curriculum at Nursery B: “*It did take a lot of time. [...] It was like trying to play catch up with other things.*” If the typical mathematics curriculum that children received was disrupted to some extent, this may have reduced learning gains.

However, in general, facilitators felt that the intervention was easy to setup and implement once children and staff were familiar with routines: “*Once we’d done it a couple of times we were quite quick at getting it all sorted and up and running.*” (Clara)

4.2.4.6 Research Question 6: Summary

In relation to Research Question 6, TAs of interview data and narrative observations indicated a number of possible factors that may have influenced the outcomes of the intervention, including: the language demands of the intervention, children’s attention on-task, the instructional level of the intervention, children’s attitude to learning and how the intervention was implemented in the setting. However, a number of contradictions were noted between facilitators, possibly indicating differences in pedagogical perspective, attitude to the intervention and nursery context.

4.2.5 Quantitative data analysis for Research Question 6

In order to further address Research Question 6 and the factors that may have affected the outcomes of the intervention, additional quantitative data analyses are presented here. First, structured observations of children’s attention during the intervention sessions are analysed through descriptive statistics. Second, correlational analyses are presented to determine whether there were any significant associations between learner’s characteristics and the gains they made during the intervention.

4.2.5.1 Data analysis of structured observations

Following each structured observation, the researcher calculated the percentage of observations that were on-task to provide an overall measure of the group's attention during the intervention sessions.

A bar chart of this data was constructed to examine any changes in children's attention and engagement over time and across different settings, see Figure 4.10.

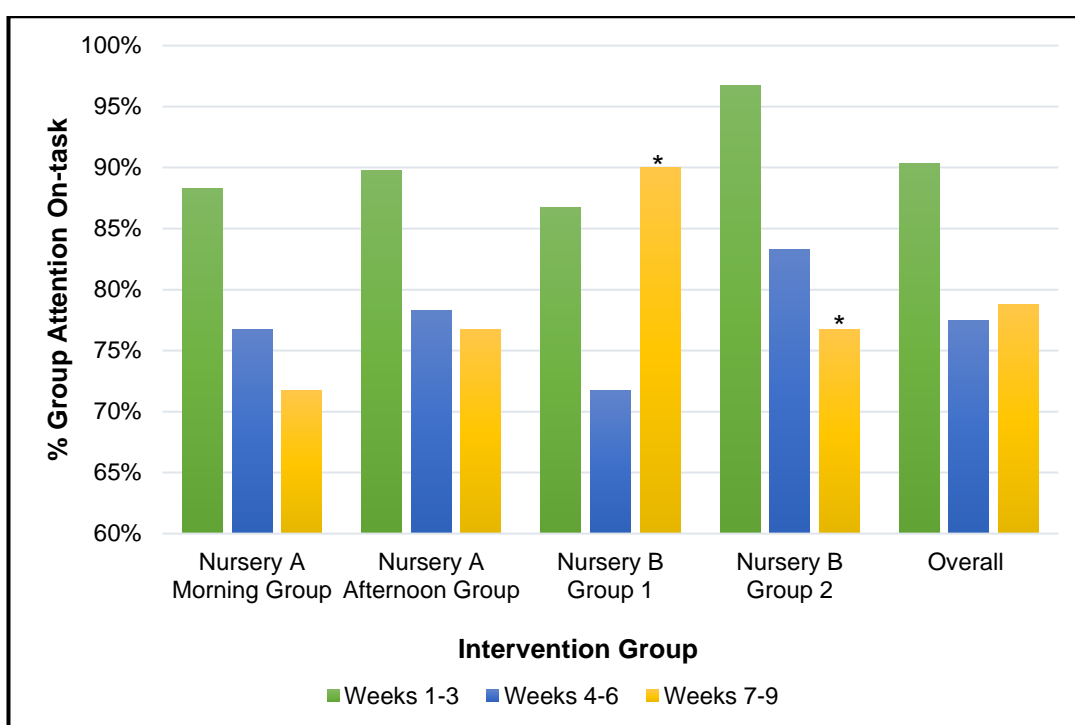


Figure 4.10: A bar chart displaying children's attention on-task during intervention sessions, by group and week of the intervention

Figure 4.10 indicates that overall children's attention on-task was highest in the first three weeks of the intervention but then decreased in the remaining weeks, indicating a possible novelty effect. This pattern was consistent for the majority of the intervention groups at each setting. However, on-task behaviour was actually highest in the last week for intervention Group 1 at Nursery B, although it is important to note though that during the last observation at Nursery B some children swapped between Groups 1 and 2 due to timetabling constraints. This might affect the comparisons between these observations (indicated with a * on the chart) and previous observations for that group. Other factors may also

affect the comparability of group attention scores over time (e.g. slight variation in facilitator, some individual child absences, day of the week).

4.2.5.2 *Analysis of associations between learner characteristics and intervention gains*

Correlational analyses were also conducted to explore whether there was any association between key characteristics of the children and the gains that they made whilst using the intervention. Learning gains were calculated for all children in the intervention group by subtracting pre- and post-test mathematics scores on the CK assessment. The researcher then determined whether these gains were associated with any of the children's pre-test scores on the mathematics assessments (ENC/CK tests), receptive language (CELF-P2) or AtL (teacher-rated 'Initiative' questionnaires from the DECA-P2). Pearson's Product Moment Correlation Coefficient (Pearson's r) was calculated to examine relationships between variables.

In order to calculate Pearson's correlation coefficients, the data must first meet the following statistical assumptions:

- The relationship between the two variable should be linear (roughly a straight line, rather than a curve).
- There should be homoscedasticity: variability in scores for variable X should be similar at all values of variable Y.
- There should be no extreme outliers.
- Data should be continuous (interval or ratio level).
- Data should be normally distributed.

(Pallant, 2016)

Scatterplots were constructed to check the linearity and homoscedasticity of the data, as well as possible outliers. Inspections of scatterplots indicated that the majority of assumptions were met across the data set. One outlying score, however, was identified on each scatterplot, relating to one particular child. These data points were therefore removed from subsequent analyses.

Subsequent Shapiro-Wilk tests were not significant for any of the variables included in the analyses, indicating that the data did not differ significantly from

normality. All data were considered to be at least interval level (see Appendix 8.16). Once outliers were removed, all necessary statistical assumptions were met to calculate Pearson's r between all variables.

Table 4.11 below presents the results of the correlational analysis for learning gains on the CK assessment by T2 and learner characteristics. Sample sizes are reported alongside the correlation coefficients and p values. It should be noted that when multiple analyses are conducted on a data set, there is a greater likelihood of identifying a 'false-positive', or Type 1 error. The α level (0.05) can be adjusted, using Bonferroni correction, to make the criteria for accepting a null hypothesis more stringent in these circumstances (Howell, 2010). Due to the small sample ($n = 18$) and low statistical power of these analyses, however, the researcher did not apply this correction as it would increase the risk of a Type 2 error (i.e. the possibility of missing a true effect).

Table 4.11: Correlational analyses between learning gains on the mathematics Curriculum Knowledge test and learner characteristics

Learner Characteristic Variable at Pre-Test	$N =$	Pearson's Correlation Co-efficient (r)	P Value
Mathematics (ENC Ability score)	18	.506	.032*
Mathematics (CK Raw score)	18	.300	.226
Receptive Language (Total Raw Score on CELF-P2)	18	.625	.006*
AtL (Initiative Raw Score on DECA-P2)	18	.164	.516
SES (IDACI Score)	16	-.380	.147

There was a statistically significant association between children's initial receptive language skills and learning gains; children with higher receptive language skills at T2 also made greater progress over the intervention period. There was also a trend towards a significant association between children's initial mathematical ability on the ENC ability test and learning gains; where

children with higher initial ENC ability scores made greater progress in CK over the intervention period ($p = .032$).

There were no statistically significant associations between any other learner characteristics and learning gains. However, Pearson's Correlational analysis also identified a significant correlation between children's progress through the graded levels of the intervention and their learning gains ($n = 18$, $r = .565$, $p = .014$). Children at Nursery A made significantly more progress through the intervention than children at Nursery B, $t(21) = 4.21$, $p < .001$.

4.2.5.3 Research Question 6: Further summary of findings

Overall, additional quantitative analysis provided further evidence that children's attention on-task declined over the course of the intervention, as reported by facilitators, and suggest a possible novelty effect. In addition, correlational analyses provided further support for the facilitators' view that children with lower language skills and lower initial attainment benefited less from the intervention.

4.2.6 Chapter summary

This chapter has presented the findings of the study in relation to each of the six research questions. Inferential statistical analyses indicated that the intervention had a significant impact on children's mathematical CK at post-test, although these gains did not remain statistically significant 5 months later. There was not a statistically significant intervention effect on any other DVs.

Additional exploratory analysis of embedded aspects of the data has identified facilitators' perceptions of the outcomes of the intervention, as well as illuminating some of the mechanisms that may underpin its efficacy.

5 Discussion

This chapter discusses the key findings of this research project in relation to previous literature and theory, considering the impact of the intervention on mathematics, receptive language and AtL, as well as factors that may have affected outcomes. Limitations in study design which might affect the interpretation of findings are also acknowledged. Finally, implications of the findings are identified, in relation to the implementation of technology in early years' mathematics education, the work of EPs and potential avenues for future research.

5.1 Research Findings and Theoretical Relevance

5.1.1 Outcomes of the intervention for mathematical attainment

The primary purpose of this study was to investigate whether a tablet-based mathematics intervention, underpinned by evidence-based principles of effective instruction (Baker et al., 2002; Doabler & Fien, 2013; Gersten et al., 2009), would improve the mathematical attainment of preschool children, aged 3-4 years (*Research Question 1*). Consistent with previous research with school-aged children (Outhwaite et al., 2017, in press), after 9 weeks, the *onebillion* intervention had a statistically significant impact on children's mathematics CK, assessed on a measure developed by previous evaluators to test topics taught directly by the intervention (Outhwaite et al., 2017). There was a small to moderate ES on this measure ($d = +0.38$), comparable to the magnitude of effects reported by Pitchford (2015) ($d = +0.35$) and Outhwaite et al. (in press) ($d = +0.31$). These quantitative findings were consistent with the views of some facilitators (*Research Question 5*), who perceived improvements in some children's mathematical knowledge/skills, particularly counting and number recognition – both elements of *number sense* thought to be foundational for later learning (Jordan et al., 2009; Jordan et al., 2012).

Contrary to experimental hypotheses, however, a statistically significant intervention effect was not identified on the ENC sub-test (Elliot & Smith, 2011),

a standardised measure of children's understanding of number, It is possible that the null result on this measure was due to low statistical power, given that a small, yet potentially still educationally significant (Higgins, Kokotsaki et al., 2012), difference between groups was indicated by the descriptive statistics (ES: $d = +.23$). The magnitude of the ES was still smaller, however, than on the CK assessment.

A number of possible factors might explain the mixed pattern of results obtained in this research, as discussed below:

5.1.1.1 Proximal and distal measures of mathematics attainment

Significant findings are more commonly reported on proximal, researcher-developed assessments rather than standardised measures across early mathematics research (Cheung & Slavin, 2016; Gersten, 2016). As noted previously, the content of the ENC sub-test does not assess all aspects of mathematics taught by the app, focusing upon children's skills in the domain of number, rather than broader aspects of mathematics, such as shape (see Appendix 8.24). In contrast, the CK assessment contains items relating to each of the intervention topics, potentially increasing sensitivity to intervention gains.

In addition, the context of questions differs to those presented in the app (e.g. story problems) and some items rely on an understanding of mathematical language not directly used within the *onebillion* intervention (e.g. 'altogether' indicates addition). In line with the Instructional Hierarchy (Alberto & Troutman, 1986; Haring & Eaton, 1978; Martens & Witt, 2004), it may be that children had not learnt skills to a level of 'generalisation', where they could transfer their learning to the new contexts presented in the ENC sub-test. This view is consistent with one teacher's perception of a lack of skill generalisation (see Section 4.2.3.2). Nevertheless, as other facilitators identified occasions where children had applied some newly acquired skills (such as counting) during other activities, the extent of generalisation may have varied across individuals and skills of varying complexity/novelty, dependent upon the fluency level achieved and opportunities for application.

5.1.1.2 *Limitations of the curriculum knowledge assessment*

Possible measurement effects on the CK assessment should also be acknowledged. First, children in the intervention group may have been at an advantage on this assessment due to greater familiarity with the format and style of questions, which are closely based upon app activities. Second, the CK assessment may be disproportionately weighted towards assessing number recognition, given that many items require number recognition alongside other mathematical skills. The intervention effect may, therefore, partially represent improvements in number recognition, rather than broader mathematical knowledge. In addition, a number of children in the sample had a low score on this test at T1 ($n = 9$ scored ≤ 3 points); consequently, the reliability of scores and sensitivity of the test at T1 may have been constrained to some extent by a 'floor effect' (Coolican, 2014). Lastly, the researcher noted that some children appeared to find it difficult to maintain attention throughout the CK assessment, due to the length of the measure. The assessment might therefore also reflect *possible* improvements in attention from using the intervention, as discussed later.

Notwithstanding these limitations, the CK assessment was selected due to its greater potential sensitivity to learning gains, and to allow comparability with findings from previous research, which have identified significant intervention effects on this measure with 4-5 year old children (Outhwaite et al., 2017).

5.1.1.3 *Differences across nursery*

It is, however, noteworthy that descriptive statistics suggested greater effects of the intervention at Nursery B on the ENC sub-test at T2, whilst there was no observable effect at Nursery A (see Figure 4.3)²⁰. Whilst this discrepancy could partially reflect differences in aspects of implementation, known to influence programme outcomes (Horner et al., 2017; Nordstrum et al., 2017), the magnitude of intervention effects appeared similar across settings on the CK assessment.

²⁰ The statistical significance of differences between setting was not tested through the use of inferential statistics due to low statistical power.

In order to explain possible differences across nursery, it is important to note that children at Nursery B had lower initial ENC ability scores than those at Nursery A. Children at Nursery A also made greater progress through the intervention, therefore accessing later topics taught within the app. The ENC sub-test may therefore have been more sensitive to capturing intervention effects on earlier developing skills at Nursery B (e.g. counting principles, number recognition to 5) than the later developing skills which may have improved at Nursery A (e.g. understanding of number over 10, early addition/subtraction skills). Whilst many of these later developing skills are assessed in the ENC sub-test (see Appendix 8.24), some are tested in unfamiliar contexts or story-based problems, requiring greater skill 'generalisation' (Haring & Eaton, 1978), or alongside other more complex skills not taught directly in the intervention.

Furthermore, due to the stopping points inherent in the standardised delivery of the ENC sub-test (see Section 3.4.5), more complex skills may not have been assessed for some children at T2 and falsely assumed 'incorrect', leading to a form of ceiling effect (Coolican, 2014) at Nursery A. Use of stopping points may also have reduced the reliability of scores obtained at T1, potentially artificially lowering some children's scores and increasing the size of observed learning gains if more items were administered at T2. Despite some correction of these effects through the use of 'ability scores' rather than raw scores, administration of all items within the ENC sub-test may have reduced the impact of possible measurement effects. The researcher was, however, conscious of the need to limit the number of test items due to ethical considerations and to reduce demands on young children's attention.

5.1.1.4 A note regarding individual differences

It is also important to recognise that there was also a high level of variability in the learning gains made by the intervention group children on both the CK assessment ($M = 8.00$, $SD = 6.46$) and the ENC sub-test ($M = 15.14$ $SD = 13.58$). Moreover, 'individual differences' in outcomes was a theme identified across facilitators' perceptions of outcomes, as noted in other early mathematics interventions (Fuchs et al., 2012; Salminen et al., 2015). These individual differences may affect the interpretation of the group level evaluation

outcomes discussed above (Pawson & Tilley, 1997). Possible factors affecting individual differences in Rtl are discussed further in Section 5.1.5.

5.1.1.5 Summary

It is likely, therefore, that a number of factors may have contributed to the discrepancy in the significance of findings between the researcher-developed CK assessment and the standardised ENC sub-test. In particular, there is need to consider whether the intervention had supported children to generalise their learning of more complex skills, as well as possible limitations to the measures employed. Nevertheless, ESs on both measures show positive benefits from supplementary use of the *onebillion* tablet-based intervention, and nested, explicit mathematics instruction (Baker et al, 2002; Doabler & Fien, 2013; Gersten et al., 2009), within the broader preschool curriculum.

5.1.2 The longitudinal impact of the intervention

A further aim of this study was to explore whether the intervention had a lasting impact on children's mathematical attainment, 'closing the gap' at school entry (*Research Question 2*). Consistent with the findings of Outhwaite et al. (2017), descriptive statistics indicated some lasting difference between groups on the CK assessment at 5 month follow-up (ES: $d = +0.21$); the size of the effect, however, had reduced from post-test (ES: $d = +0.38$) and was no longer statistically significant, restricting the conclusions that can be drawn. The reduction in effect appeared to be due to a slower rate of learning for children at Nursery A following withdrawal of the intervention, allowing children in the control group to 'catch-up'.

Despite predictions from theories of cumulative learning (Gagné, 1968), longitudinal fade-out is common in early mathematics intervention research (Bailey et al., 2016). As previously recognised, fade-out effects may be explained by a number of factors. First, in line with predictions from the Instructional Hierarchy (Alberto & Troutman, 1986; Haring & Eaton, 1978; Martens & Witt, 2004), it is possible that some newly acquired knowledge, particularly more advanced content learnt by children at Nursery A, was not learnt fluently by the end of the intervention, and was therefore perhaps forgotten or not yet sufficiently consolidated to support later learning (Bailey et

al., 2016); this might account for an apparently slower rate of learning in the experimental group post-intervention. It should also be borne in mind that, by T3, children in this study had made a transition to a new class following the long summer break, potentially affecting skill maintenance.

Moreover, consistent with the findings of Bailey et al. (2016), the intervention may not have had a significant impact on more stable factors that continued to influence children's progress over time, such as cognitive skills, language (as discussed below), motivation and opportunities for mathematical learning at home, reducing their learning rate. It is, however, important to note that these effects were not apparent at Nursery B; the sample size was, nevertheless, small for this group by T3 due to attrition ($n = 5$), limiting the conclusions that can be made by nursery. The impact of the intervention on broader developmental factors is discussed further below.

Finally, the mathematics curriculum that intervention children received in their new classes post-intervention, particularly for higher-attaining children, may not have supported children's progress in more advanced topics assessed on the CK assessment (Bailey et al., 2016). Teachers may not have taught these skills, perhaps due to a lack of awareness of children's attainment following the intervention, or a need to focus on more basic skills for those children in the class who had not received the intervention. It is also important to note that some skills assessed in the CK measure would perhaps not typically be taught until Year 1 (e.g. recognising odd/even numbers, symmetry) (DfE, 2013); teachers may have been focusing upon *mastery* learning of earlier content (NCETM, 2014), or other aspects of mathematics not directly captured by the CK assessment. Nevertheless, follow-through training for new class teachers in the content of the programme and children's level of progress may be valuable in securing on-going learning trajectories, consistent with approaches found to be effective by Clements et al. (2013) in a longitudinal evaluation of the *Building Blocks* early mathematics intervention.

Finally, measurement effects may also have played a role, given that delayed post-testing occurred during school Christmas activities and preparations. This may have affected children's focus on-task and the reliability of assessments,

accounting for increased variability, and potentially obscuring any group differences in attainment.

These hypothesised explanations are, however, not mutually exclusive. It is also important to recognise that the non-significant difference between groups at post-test (ES: $d = +0.21$) may still be educationally meaningful (Higgins Higgins, Kokotsaki et al., 2012); further replication may be beneficial with a larger sample size.

5.1.3 Outcomes of the intervention for receptive language skills

This research also aimed to evaluate the impact of the *onebillion* intervention on other areas of broader aspects of children's development, including early language skills (*Research Question 3*). It was predicted that the *onebillion* intervention might support the development of children's receptive language given that (a) children are provided with increased opportunities to listen to instructions from a virtual teacher and (b) the software provides explicit instruction in semantic concepts and vocabulary related to mathematics. In this study, however, no significant intervention effect was identified on children's receptive language skills, as measured by the CELF-P2 (Wigg et al., 2006).

On the one hand, this null result may be due to the fact that the instructions and specific mathematics language tuition provided through the *onebillion* app software were not developmentally attuned for some children, and that they were not able to learn within their ZPD (Vygotsky, 1978). As discussed below, the language demands of the intervention may have affected some children's progress. Use of the app might therefore be more beneficial for the language development of children who have sufficient prerequisite skills.

It is also, however, important to consider the positioning of the *onebillion* intervention within the wider mathematics curriculum that the children were receiving. Sarama et al. (2012) noted a positive impact on oral language skills in the *Building Blocks* mathematics intervention. In that programme there is an emphasis on children learning mathematical language through discussion, alongside their learning through technology; for example, in small group activities, teachers ask children to justify their mathematical reasoning by

responding to questions such as ‘*How do you know?*’ and ‘*Why?*’, and engaging them in discussion about the meaning of particular mathematical words. During these activities, children may therefore develop a more precise grasp of semantic concepts, as well as improving their syntactic skills from listening to and explaining their reasoning (Sarama et al., 2012). Greater pedagogical emphasis on mathematical ‘talk’ prior to, or alongside, use of the app software, may therefore be beneficial.

Although no significant effect was identified in this study, it is possible that use of the apps held some benefits for language development that were not identified due to measurement sensitivity. As the CELF-P2 is a standardised measure of general receptive language skills, including syntax and a wide range of semantic concepts (see Section 3.4.5), some of the specific mathematics vocabulary taught within the *onebillion* intervention was not assessed through these sub-tests. Further research could therefore more closely evaluate the impact of the *onebillion* intervention on the mathematical language supported through the app software.

5.1.4 Outcomes of the intervention for approaches to learning

It was also hypothesised that using the *onebillion* tablet intervention might support the development of more positive AtL (*Research Question 4*), given that it would provide children with opportunities to persist in learning, problem-solve and develop confidence through receipt of positive feedback. Statistical analysis, however, indicated that the *onebillion* intervention did not have a significant effect on children’s AtL, as reported on the teacher-rated ‘Initiative’ scale of the DECA-P2 (LeBuffe & Naglieri, 2012).

There are a number of potential considerations in relation to this finding. First, AtL is thought to be promoted when children show high levels of task engagement (Williford et al., 2013); however, qualitative analyses indicated that attention-on task may have been compromised for some children, particularly when they did not understand the task (see Section 5.1.5). Adaptations to the pedagogical features of the app software, such as additional layers of instructional scaffolding (Vygotsky, 1978; Wood et al., 1976) for children experiencing difficulties, could enable all children to experience greater success

and facilitate problem-solving, thereby promoting persistence, independence and confidence in learning. In addition, given the potential role of teachers in promoting AtL (Ansari & Gershoff, 2015; Williford et al., 2013), explicit training for facilitators in pedagogical approaches to support AtL during the use of technology may have been beneficial, including commenting on and praising positive learning approaches, and helping children to reflect on their actions to support problem-solving and persistence (e.g. “*How else could you try that?*”) (Chen & McNamee, 2011).

It is also important to consider whether the type of measure used in this study, a teacher-report scale, was sensitive to all nuanced changes in AtL over time. Teachers may not have been aware of changes in all children’s behaviour across a short period of time, and perhaps at times ratings were influenced by teachers’ expectations of behaviour, rather than direct observations. At Nursery A, where there was a delay in the return of class-teacher questionnaires until after the six-week summer holiday period, the reliability of teacher assessments may also have been weakened. These effects may have constrained the possibility of identifying a significant post-test difference between groups.

The effect of individual differences on outcomes and the multi-component nature of AtL should also be recognised. For example, some children appeared to have developed greater self-confidence in learning following the intervention (see facilitator interviews/researcher observations). Moreover, consistent with the findings of Pitchford and Outhwaite et al. (in prep.) that the *onebillion* apps might promote attention, two facilitators also noted improvements in some children’s engagement on-task. It is therefore possible that some components of AtL improved for some of the children in the study, but that these effects were not captured when analysing group level effects.

5.1.5 Factors affecting the outcomes of the intervention

Additional aspects of data collection/analysis embedded within this study allow further illumination of the mechanisms underpinning the efficacy of the *onebillion* intervention and features of the intervention which may have affected the outcomes considered above (*Research Question 6*). A number of data sources were used to address these factors, including narrative observations,

semi-structured facilitator interviews, structured observations of children's attention on-task and correlational analyses.

Based upon analyses of this data, a model of the underpinning mechanisms of the *onebillion* intervention in this study is proposed (see Figure 5.1). On the left, are key characteristics of children which may have affected their learning experience during the intervention. In the centre, are five inter-related factors that may have affected children's individual learning experience, derived from the five main themes identified from TA of qualitative data (see Section 4.2.4). The final outcomes of the intervention are identified on the right of the model. Particular features of the intervention that may have affected children's learning experience are then identified in relation to four key areas:

- features of the software/hardware,
- the role of the facilitator,
- the role of peers, and
- broader contextual factors.

The model is discussed in detail below, in relation to key supporting data, psychological theory and previous research.

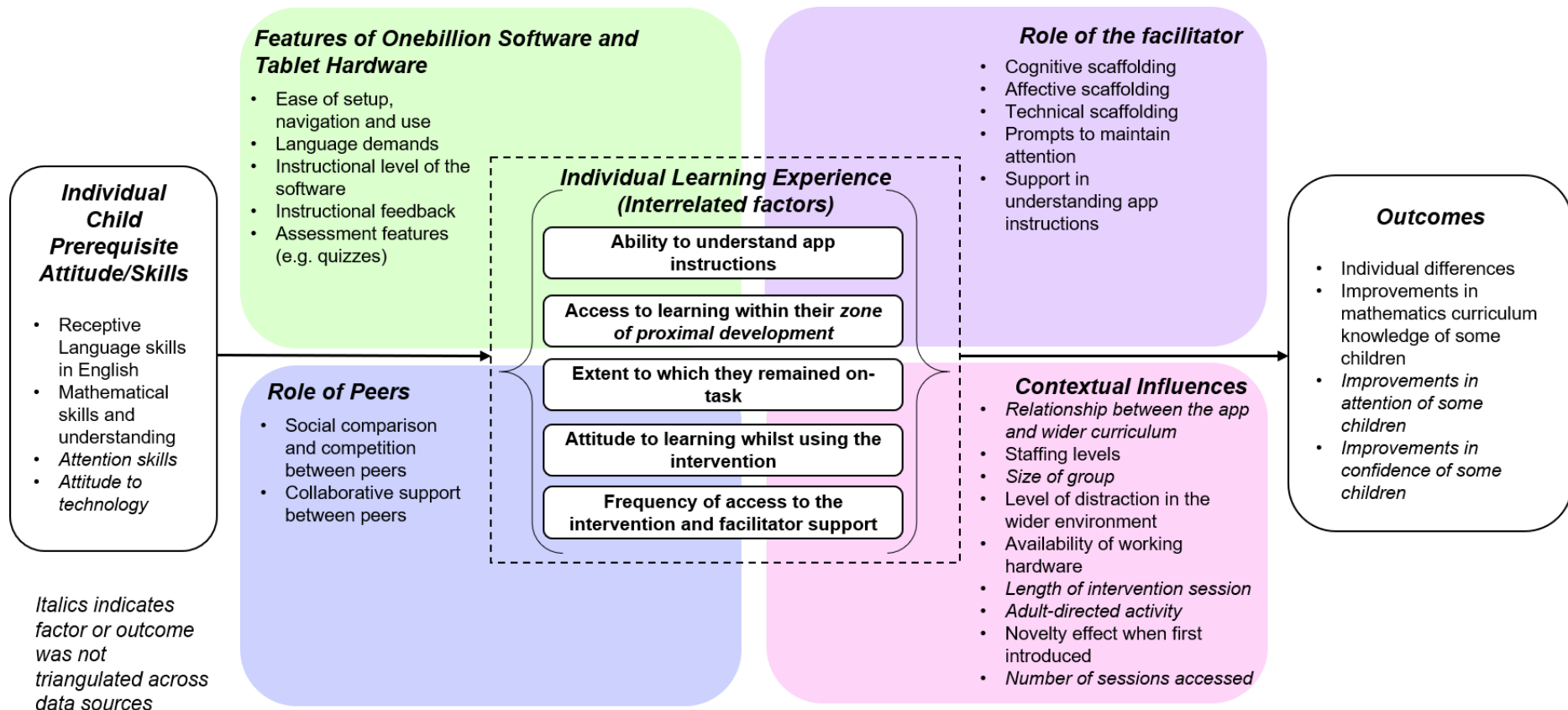


Figure 5.1: A model of the mechanisms underpinning the efficacy of the onebillion tablet intervention

5.1.5.1 *Child's ability to understand app instructions*

Consistent with the view that early language skills provide a gateway to mathematical learning (Praet et al., 2013), both correlational analyses and facilitator interviews indicated that children with lower language proficiency in English appeared to benefit less from the intervention, perhaps finding it difficult to independently understand app instructions and mathematical vocabulary. These findings, however, contradict those of Outhwaite et al. (2017) who did not find a significant relationship between 4-5 year old children's receptive vocabulary skills and learning gains (Study 2), or a significant impact of EAL status on outcomes (Study 1). Discrepancies across studies most likely reflect the younger age of the preschool children in the present study, or other sampling characteristics, given the small scale of both studies.

5.1.5.2 *Extent to which child remained on-task*

Consistent with the observational findings of Pitchford et al. (2018), facilitators felt that children who had difficulty sustaining attention-on task benefited less from the *onebillion* intervention. They also felt that the 15 minute duration of the intervention was too long for some children. Although the duration of sessions was reduced to 15 minutes from the 30 minute sessions delivered in past studies (Outhwaite et al., in press; Outhwaite et al., 2017), shorter sessions, spaced over time, may be more effective for preschool children, employing principles of *distributed practice* (Cepeda et al., 2009; Son & Simon, 2012), thought to be more effective for young learners and improving retention (Seabrook, Brown & Solity, 2005; Shapiro & Solity, 2008).

To some extent, children's level of on-task behaviour may also have been affected by distractions in the wider environment (e.g. noise, activities of other children), particularly at Nursery B, consistent with the notion that implementation context can affect outcomes (Horner et al., 2017; Nordstrum et al., 2017). Some facilitators considered, however, that particular features of the *onebillion* intervention may have supported children's attention, such as task variety and the adult-led nature of the activity; higher levels of teacher instruction have been found to promote positive leaning behaviours, including task engagement, in young children (Ansari & Gershoff, 2015). Other

motivational features of the software, noted later, may also have supported on-task behaviour.

The extent to which children sustained attention may, however, have been related to the level of instructional challenge: the researcher observed children off-task on a number of occasions after they had been unsuccessful in completing activities, suggesting that they did not understand the activity and may have been engaging in task avoidance (Roberts, Marshall, Nelson, & Albers, 2001). At times, greater instructional scaffolding from facilitators may have helped to promote on-task behaviour more effectively, as discussed below.

5.1.5.3 Child's ability to access learning within their 'zone of proximal development'

Vygotsky (1978) emphasised the importance of children learning new skills within their ZPD (Section 2.4.4) and with the appropriate level of cognitive scaffolding. During the *onebillion* intervention, facilitators felt that children's learning was scaffolded by particular features of the app software, including the 'real world' objects in the app; familiarity is thought to enhance learning with manipulatives (Carbonneau et al., 2013). They also perceived benefits from the visual elements of the app, potentially enhancing multi-sensory learning (Pavio, 1986; Carr, 2012) and providing 'iconic' representations to support abstract and symbolic understanding of concepts, such as number recognition (Bruner, 1966). The researcher observed many children moving successfully and independently through app activities, suggesting that features of explicit, direct instruction embedded within the technology were appropriately scaffolding learning, including modelling from the virtual teacher (Vygotsky, 1978), instantaneous feedback (Haake et al., 2015; Henderson & Yeow, 2012) and a staged, task-sliced curriculum (Magliaro et al., 2005).

Qualitative data indicated that the social context of the intervention also appeared to provide additional cognitive scaffolding for learners, both from the support provided by facilitators (Yelland & Masters, 2007), and, in some cases, from peers; potential benefits from peer collaboration in early childhood education have been highlighted in previous studies, particularly for lower

attaining children (e.g. Park & Lee, 2015; Fawcett & Garton, 2005; Fuchs, Fuchs & Karns, 2001).

The results of this study, however, indicate some variation in the extent to which children were learning within their ZPD; facilitators felt that lower attaining children benefited less from the intervention and children with lower initial early number skills made fewer gains in mathematics CK over time. These results indicate a potential Matthew effect, where those children with the strongest initial skills make the greatest gains, whilst those with weaker skills fall further behind (Stanovich, 1986; Hindman, Erhart & Wasik, 2012). Findings contrast with previous evaluations of the *onebillion* intervention, which found particular benefits for lower attaining pupils (Outhwaite et al., in press; Outhwaite et al., 2017). Again, the younger age of the pupils in this study might explain these discrepancies, perhaps suggesting that there is a particular ‘developmental window’ in which the intervention is most beneficial.

Incorporating additional graded instructional scaffolding within the app software, a key feature of another promising tablet intervention, *Math Shelf* (Schacter & Jo, 2016, 2017; Schacter et al., 2016), might support children initially experiencing difficulties, achieving greater personalisation in the delivery of instruction (Gifford & Rockliffe, 2012; Holmes & Dowker, 2013). Moreover, whilst facilitators reported that the app helped them to assess where children were in the curriculum and inform next steps, nesting the app more closely within children’s wider mathematics curricula, as suggested by some facilitators, might enhance outcomes, especially in supporting fluency with newly acquired skills and generalisation beyond the intervention (Haring & Eaton, 1978). If some adult-led teaching is provided first, it may also enable children working at a more *enactive* level of skill development (Bruner, 1966) to have more concrete experiences, before progressing to a more *iconic* level of understanding, required by many of the intervention activities.

5.1.5.4 Child’s attitude to learning whilst using the intervention

Despite a null association between teacher-rated AtL and learning gains for the intervention group, affective factors may still have played a role in children’s

learning experience; 'Attitude to learning', was identified as a key factor that may have affected outcomes from TA of facilitator interviews/observations.

Children's attitude to learning whilst using the apps appeared to be influenced by the feedback that they received. For children who were experiencing regular success, positive feedback from the app software (e.g. certificates after app quizzes) and praise from teaching staff may have provided positive reinforcement (Skinner, 1953), motivating children to progress through the intervention and influence outcomes, given that progress was significantly related to learning gains. Research indicates that receiving positive feedback can also develop feelings of self-efficacy and competence (Schunk, 1983), which, in turn, may have increased children's intrinsic motivation (Deci & Ryan, 2000). Conversely, for children who received higher levels of negative feedback, there may have been declines in motivation, as noted by facilitators, potentially reducing progress and learning gains. Avoidance of negative feedback might also have contributed to some hesitancy in responding, which was observed by the researcher during app activities, despite affective scaffolding from facilitators (Yelland & Masters, 2007).

Peer group dynamics also appeared to be another important influence on children's attitude to learning during the intervention, given that both facilitators and the researcher noted that an element of 'competition' appeared to be motivating children, with children comparing their relative position through app activities. According to social identity theory (SIT) (Tajfel 1978; Tajfel & Turner, 1979), individuals are internally motivated to join higher status social groups, (i.e. 'the highest level in the app') in order to enhance their own self-esteem. Even young children are thought to make social comparisons in line with the predictions of SIT (Nesdale & Flessner, 2001; Yee & Brown, 1992).

Finally, an apparent novelty effect was also observed by some facilitators, triangulated through structured observation which showed a decline in attention on-task for most groups after the first three weeks of the 9 week intervention. Again, this study suggests that shorter intervention periods may be advisable for nursery children, following principles of distributed practice. Further research might also explore other factors that some facilitators felt may have affected

children's motivation (e.g. children's prior attitude to technology, level of choice about when to use the intervention).

5.1.5.5 Frequency of access to the intervention and facilitator support

Qualitative data also indicated that a number of broader implementation factors, thought to influence findings from the scale-up of evidence-based practice research studies and the longer term use of interventions (Horner et al. 2017), which may have affected children's ease of access to the *onebillion* software and the availability of staff to provide additional support, including technical difficulties with hardware, group size, staffing levels and missed sessions, indicating that these factors need careful consideration in the setup of the programme. Nevertheless, apart from a 'bug' in one of the software activities, facilitators felt that the intervention was easy to implement and the software was user-friendly for children, potentially supporting their progress and, therefore, learning gains.

5.1.5.6 Summary

The mechanisms underpinning the efficacy of the *onebillion* intervention are somewhat tentatively proposed given the exploratory nature of this aspect of the research design and the small scale of the study. Nevertheless, they provide insight into possible reasons for differences in Rtl between children, and may have important implications, not only for the future development and use of the *onebillion* software, but also for the wider implementation of educational technology in early years education, as discussed in Section 5.3.

5.2 Methodological Limitations

5.2.1 Limitations of the quasi-experimental design

Although steps were taken to address possible threats to reliability and validity within the main quasi-experimental design used in this study (see Section 3.4.7), a number of key limitations should be acknowledged.

5.2.1.1 Internal validity

First, it is uncertain whether the gains made by the intervention group in mathematics CK were due to the additional time spent on maths-based activity,

rather than the *onebillion* intervention *per se*. Previous research has included comparison groups, who received adult-led tuition in topics similar to those on the *onebillion* app (Outhwaite et al., in press) or even other maths tablet apps (Schacter et al., 2016, 2017). Inclusion of a comparison group in the current study was not feasible due to class staffing levels/sample size, but would have been valuable in determining the relative impact of the *onebillion* intervention compared to adult-led instruction or other interventions. It would have also reduced the possibility that any group differences were, at least in part, due to a potential Hawthorne effect (Roethlisberger & Dickson, 1939), from children receiving ‘special’ attention.

As noted previously, RCTs are typically thought to be the gold standard in educational research (Slavin, 2002; Cook, 2009), given that they eliminate selection bias and enhance the internal validity of the study. The sample size of the present research precluded the use of an RCT due to concerns that randomisation may not lead to an even spread of ability across both conditions (Robson, 2011); children were instead allocated to condition using an ability matching procedure. There is, however, a possibility that some differences in unknown confounding variables remained between the groups. Researcher bias in measurement of ENC ability, used for initial matching, and the reliability of assessments conducted on this measure may also have influenced group allocation. Nonetheless, the researcher aimed to administer this assessment according to standardised test instructions and without prior knowledge of the children.

In order to control for differences in the wider environment and teaching that children received, experimental and control children were selected from within the same classes. As discussed in Section 5.1.1, this may have led to some compensatory additional teaching being provided to children in the control group by some facilitators, or there may have been some diffusion of treatment effects, from children in the control group accidentally accessing the *onebillion* intervention, perhaps at home, or teaching staff altering the delivery of the whole class mathematics curriculum from experience in facilitating the intervention. However, during fidelity checks, there were no teacher reports of

children in the control group accessing the intervention, or children from either group accessing the intervention between T2 and T3.

Lastly, as acknowledged above, the sample size of the present study may have limited statistical power to detect a statistically significant intervention effect on key DVs. Due to practical constraints in sampling and recruitment, the desired sample size ($n = 52$) was not reached and therefore replication of this study would be valuable with a larger sample.

5.2.1.2 *External validity*

The small scale of the study may also affect the generalisability of these findings. However, demographic details of participants and nurseries are provided to support interpretation of the possible relevance of the results in other contexts; findings from Nursery B, for example, may have stronger generalisability to nursery settings where there is a greater proportion of children with EAL or lower SES than the national average.

Moreover, there may be limitations to the ecological validity of the research. The researcher approached schools in this study to use the *onebillion* intervention for the first time for the purposes of evaluation, and the researcher carefully controlled its implementation by providing training, allocating children to condition, monitoring fidelity over time and providing ongoing facilitator support. These processes may not reflect the more *natural* implementation of the intervention by schools, perhaps reducing the applicability of the research to 'real world' contexts. In particular, the researcher recognises that there may have been some disruption to children's typical curriculum when the intervention was introduced, see Section 4.2.4, which may have reduced its impact (Mertens, 2015).

Furthermore, it is likely that, to some extent, the effects reported here were influenced by the facilitators' approach and were specific to the way that the intervention was implemented in the nursery context, as well as the extent to which children and facilitators had previous experience with tablets in the classroom (see Section 3.4.4). By exploring the potential mechanisms underpinning the efficacy of the intervention, as well as 'thick' description of participants, the researcher has endeavoured to provide sufficient detail for the

reader to make judgements about the applicability of these findings to new contexts (Green et al., 2015).

5.2.2 Limitations of embedded aspects of the design

5.2.2.1 Qualitative data

Limitations to the quality and trustworthiness of the qualitative data are discussed here according to the five core areas highlighted by Guba & Lincoln (1989): credibility, transferability, dependability, confirmability and authenticity (see Section 4.2.1).

The researcher aimed to enhance the credibility and authenticity of the TA of semi-structured interviews by sharing transcripts with facilitators by post. Whilst facilitators agreed that the transcripts were accurate, it is uncertain whether they were read closely. Member checks were possible with only three of the facilitators; it is also possible that due to social desirability bias they may have felt it difficult to disagree with the analysis presented by the researcher. As the researcher conducted the narrative observations herself, the credibility of these observations could not be externally established. However, the main themes identified from these observations triangulated with the themes identified from analysis of interviews.

As previously discussed, the research was conducted on a small scale. Interviews were only conducted with four facilitators and across seven observations, potentially reducing the transferability of findings to other contexts.

In order to enhance the dependability of this research, details are provided of the key focuses of the narrative observations and questions which guided the semi-structured interviews (see Appendix 8.13). Observations aimed to produce data that were neutral, yet any observation data are produced through the selective filter of the researcher-observer (Robson, 2011). Those aspects that the researcher felt were particularly relevant in capturing children's experiences and which were within the limits of their attention may have been noted more readily; additional research filming intervention sessions may be beneficial in

further replication of this research, not only in order to enhance dependability, but also to capture more fine-grained detail within the intervention sessions.

The researcher strived to ensure that themes and interpretations were supported by the data, conducting inter-rater checks to support the confirmability of the findings. It is likely, however, that the researcher's own positioning may have influenced the interpretations drawn (see Section 4.2.1). Moreover, the views expressed by facilitators may also have been influenced, either positively or negatively, by facilitators' attitudes towards the intervention and their own investment in running the intervention over time. Whilst some more negative viewpoints were expressed about the intervention, facilitators may also have felt a degree of social desirability bias in presenting the intervention favourably to the researcher.

5.2.2.2 Quantitative data

Consideration of the reliability and validity of the additional embedded quantitative data analysis, although more exploratory and tentative in nature than the main experimental findings, is also needed.

First, the reliability of the structured observations of children's attention on-task was not established through inter-observer checks, although the researcher did take steps to adapt the observation protocol to support greater reliability (see Section 3.5.4).

Second, the correlational analyses conducted to explore relationships between children's characteristics at pre-test and learning gains made over time do not provide direct evidence of causality. For example, whilst a significant relationship was identified between children's pre-test receptive language skills and intervention gains in mathematics, a third factor related to language skills, such as attention, may have accounted for this relationship over time. The statistical power of these analyses was also particularly constrained given that children in the intervention group only were entered into the analyses. Experimental research using sub-group analysis to explore the differential effects of the intervention for different groups of children may therefore be beneficial.

5.2.3 Ethical considerations

Whilst this research was designed with reference to appropriate guidance and received ethical approval from the UoN Ethics Committee (see Section 3.6), a number of further ethical issues arose which may warrant further consideration in similar studies.

First, although parents were invited to attend meetings at the school to inform them about the purpose and nature of the study, some parents were unable to attend. Moreover, on consent forms some parents indicated that they had not had opportunity to ask questions, prompting the researcher to send a further letter to these parents providing contact details and an offer for further discussion. However, as the researcher received no further contact, it is possible that not all parents were fully informed prior to giving permission for their children to participate in the study.

Second, facilitators identified that some children experienced frustration, or became disheartened, when they did not make progress through the intervention. During nursery visits, the researcher discussed these concerns with facilitators and suggested that they provide additional pedagogical and/or affective support to children to enable them to experience success. Nevertheless, careful consideration of the suitability of the intervention for each individual, and the level of additional adult support they may require, should be carefully monitored if the *onebillion* intervention is used in the future.

Facilitators also identified that some children in the control group were keen to use the intervention, and may have been disappointed that they could not use the tablet technology. The researcher aimed to address these concerns during the intervention period by asking teaching staff to ensure that all children would have regular access to tablets at other times, and was assured that this was the case through direct observation and in conversation during fidelity checks. The researcher also initially intended that all children in the control group would access the intervention after the end of the experimental period (as a wait-list control); however, this was not feasible in either school due to a combination of practical constraints (e.g. staffing levels in the Foundation Stage 2 class,

difficulties with the availability of hardware). It would, however, be beneficial in further research studies, particularly if there are lasting benefits for attainment.

Finally, facilitators at Nursery B commented that the introduction of the intervention had caused some disruption to their teaching of the rest of the curriculum and had restricted the availability of iPad tablets for other children. Additional discussions with staff regarding timetabling at the start of the intervention period may have been valuable, although staff initially felt that the frequency/length of the intervention would be manageable, and at no point did they express a wish to withdraw from the study.

5.3 Research Implications

5.3.1 Implications for the use of tablet technology within early years education

Notwithstanding the methodological considerations noted above, the findings of the present research indicate that the *onebillion* tablet intervention may have some benefits for the mathematical attainment of 3-4 year old children and could be a valuable supplementary teaching tool in the early years classroom. In light of the effects of individual difference observed in the present study, teaching staff should, however, ensure that children using the intervention have the pre-requisite skills needed to experience success (e.g. early number skills, language, attention) and that the effects of any intervention are monitored carefully, both in the short and long-term, to ensure that they are effective for all. For some children use of the *onebillion* intervention could be more beneficial once they have already started school, as previous research has indicated more universal and sustained benefits for children in Foundation Stage 2 (Outhwaite et al., in press, 2017).

A number of potential implications for how technology-based interventions are implemented within early years classroom are also indicated. In particular, this study has revealed the importance of the social context whilst children use technology, including adult cognitive, technical and affective scaffolding, as well as peer group dynamics. Given the potentially critical role of facilitators in mediating children's learning experience through technology, additional training

and guidance may be valuable for educators prior to use of educational technology. This support should focus not only on technical aspects of implementation but also the most appropriate pedagogical approaches to foster early learning and development (Higgins, Xiao et al., 2012). Drawing upon the findings of this study, recommended guidance for nursery facilitators might include a number of considerations, as noted above, such as:

- use of Vygotskian approaches to provide graded scaffolds and reduce unfocused time spent on the app;
- fostering positive AtL through modelling, praise and questioning;
- discussion of mathematical language, during/prior to use of the app;
- timetabling short, spaced intervention sessions, following principles of distributed practice;
- establishing links between the app and the wider curriculum to support generalisation;
- securing appropriate staffing levels and availability of hardware; and
- restricting group size (potentially $n \leq 5$) to ensure that nursery children can be provided with appropriate support, including more individualised instruction for children experiencing difficulty.

5.3.2 Implications for the future development of tablet-based early years mathematics interventions

Findings from this research also have important implications for the future development of the *onebillion* software, and similar tablet-based interventions.

First, as noted, the app software itself could be improved by adding additional instructional scaffolding and support to learners based on their responses (Schacter & Jo, 2016, 2017; Schacter al., 2016), such that children who are having initial difficulty are provided with additional modelling to succeed. This support may benefit children who are unsure of how to complete app activities, reducing meaningless trial and error responding, and lowering demands on facilitators.

Second, given that the researcher observed many children repeatedly attempting the app quizzes in order to succeed, adaptations to app software so

that children are directed for further practice in relevant areas following app quizzes may be beneficial to ensure greater accuracy/fluency in learning to support success (Haring & Eaton, 1978). Embedding monitoring tools within the technology, available in Malawi (Pitchford et al., 2018), may also enable facilitators to check children's progress more easily and could help them to target pedagogical support more effectively to particular children.

Lastly, the app software might be further developed to enhance the use of virtual manipulatives and practical experiences of counting that children had during the early stages of app use (e.g. moving objects into a separate area to count them, rather than counting only static items). In line with the CPA approach (Gifford et al., 2015), these adaptations could enhance outcomes for children who require greater concrete scaffolding of particular skills.

5.3.3 Implications for the professional work of educational psychologists

A key professional role for EPs is the dissemination of the findings of research and the support of evidence-based practice in educational settings (Birch, Frederickson & Miller, 2015). This study holds implications for the guidance EPs may provide to schools/nurseries about the use of technology in early mathematics education. In particular, EPs can be made aware of potential barriers to children's response to tablet-based interventions, as identified in this study, supporting staff in the most effective implementation of tablet technology, and developing their theoretical understanding of appropriate pedagogical approaches.

In order to inform early intervention, EPs are often involved in conducting assessments with young children prior to school entry. Nevertheless, at present, there are relatively few tools available to support assessments with preschool children and to monitor the impact of interventions, particularly in the area of mathematics and AtL, as noted in Section 3.4.5. EPs might therefore have a role in further developing assessment measures for this age range.

This study has also revealed how a variety of influences may affect children's attitude to learning during the intervention. In particular, the impact of peer group factors on response to intervention was an unexpected outcome from this

study, which it may be important for EPs to consider when they are monitoring the effects of group-based interventions. This focus upon social factors also leads towards the question of the individual child's perception of their learning and environment, and therefore towards the ways in which EPs further raise young children's voice during their practice and in research, perhaps through videos, picture-based sorting activities and focus groups (Johnson, Hart, & Colwell, 2014; Merewether & Fleet, 2014), alongside careful observation.

A number of directions for further research arose from the outcomes of this study, as discussed in the next section; as scientist-practitioners (Birch, Frederickson & Miller, 2015), EPs are well-positioned to conduct further research in this area to support more effective educational practice, or to assist schools in evaluating their own intervention approaches, contributing to additional practice-based evidence within early mathematics education (Fox, 2011).

5.3.4 Directions for future research

The findings of this study also raise implications for further research into the *onebillion* tablet intervention and other technology-based mathematics interventions in the early years.

Additional research is required to further elucidate the various factors which might affect children's Rtl, aligning with the 'implementation science' movement (Horner et al., 2017; Nordstrum et al., 2017). It may, for example, be valuable to quantitatively evaluate whether possible influences noted by facilitators, such as preschool children's prior experience with technology, impacted upon learning outcomes. Further research might also further explore the most effective pedagogical use of the apps in preschool settings, drawing upon the factors identified above: for example, studies might evaluate whether adult-led instruction of app topics, incorporating additional maths 'talk', prior to exposure to the software might enhance outcomes, perhaps using a group study with children allocated to a 'linked curriculum', 'standard app use' or 'control condition'. A similar group study might be undertaken to compare *distributed* and *massed* practice with the app, considering if this leads to better learning outcomes and retention over time.

Facilitators felt that preschool children with poorer English language skills, including children with EAL, did not make as much as progress from using the *onebillion* tablet intervention. Exploration of whether the outcomes of the *onebillion* tablet intervention can be enhanced when children use the apps in their home language is warranted. Research suggests that education in the early years is most effective in the home language, in order to support future linguistic and cognitive development (Ball, 2011). Using educational technology might overcome some of the practical difficulties in providing sufficient home language support to children, particularly given the diversity of languages spoken within many classrooms in the UK (DfE, 2016). The researcher is aware of recent research, as yet unpublished, exploring the impact of the *onebillion* intervention in Brazil in both first and second languages with 5-6 year old children (Outhwaite, Neves, Gulliford & Pitchford, in prep.). Further research might be beneficial in a UK context and with younger, preschool children.

There was some evidence from this research that children provided collaborative support to one another whilst using the *onebillion* intervention. Given the potential benefits of collaborative learning in early education, noted in previous research (e.g. Park & Lee, 2015; Fawcett & Garton, 2005; Fuchs, Fuchs & Karns, 2001), quantitative studies might explore whether enabling children to use mathematics tablet interventions to problem solve collaboratively and to discuss their learning would be beneficial for attainment, potentially using split head-phone sets.

Given that facilitators noted improvements in some children's self-confidence and attention, further in-depth case-study research might be valuable in capturing change over time in aspects of AtL for particular children in more depth, perhaps incorporating children's own voice, using techniques outlined in Section 5.3.3, to explore nursery children's perceptions of the intervention and their attitude to mathematics over time.

Lastly, the findings of this study suggested that the *onebillion* tablet intervention may have been less beneficial for lower attaining children. Further research might explore whether it could, however, be an effective intervention for older children experiencing particular difficulties with learning mathematics or with

particular types of additional need. Research conducted in Malawi indicates that children with SEND made progress in their mathematical learning whilst using the *onebillion* intervention (Pitchford et al., 2018); however, controlled studies in the UK may be beneficial for children with additional needs. Use of single case experimental design methodology may be valuable in this area of research, given the likely heterogeneity within this group of learners.

5.4 Additional Reflections

Whilst conducting this research, the researcher has reflected on many of the challenges of conducting 'real world' research and a controlled study within the day-to-day activity of nursery classrooms. Negotiating the requirements of the research and ensuring commitment from teaching staff, was particularly important during the sampling and recruitment phases of the research. The researcher noted the importance of providing adequate training to facilitators at the outset, whilst also acknowledging their need for ongoing support, especially in solving technical and ethical difficulties encountered.

The researcher was also struck by the range of individual difference amongst the preschool children participating in this study and the potential impact of these differences on outcomes. In order to be effective, early mathematical intervention should therefore be flexible in meeting the specific needs of each individual (Gifford & Rockcliffe, 2012; Holmes & Dowker, 2013). Nevertheless, the majority of children appeared to respond well to the use of tablet technology, with very few technical difficulties in navigating and responding. Tablets may, therefore, be an accessible and effective mode of delivery for early intervention across many domains and an area which warrants further research.

A further reflection arising from the conduct of this study is the importance of careful monitoring of interventions by teaching staff, potentially supported by EPs. In particular, it was valuable to consider the more affective impact of a learning intervention on young children, recognising that children's levels of motivation, position within their peer group and feelings of self-efficacy are

potential contributors to outcomes. Careful monitoring of these factors alongside academic progress may therefore help to ensure the success of an intervention.

6 Conclusions

The main aim of the current mixed-methods study was to evaluate the efficacy of a 9-week tablet intervention on 3-4 year old children's mathematical knowledge and skills. Quantitative results showed that at post-test, children accessing the *onebillion* intervention on a daily basis had significantly higher mathematical CK, assessed by a researcher-developed measure, than children in a TAU control group. Children in the intervention group also scored more highly, on average, at post-test than control children on a standardised measure of early number concepts, but differences were not statistically significant. Differences across measures are likely to reflect test sensitivity and measurement effects, as well as the need for greater skill generalisation on standardised, more distal, measures of mathematics. Notwithstanding some methodological limitations, these results suggest there may be benefits from supplementary, explicit instruction in mathematics for young children, delivered through the medium of tablet technology (Magliaro et al., 2005).

Addressing a gap identified in the current literature, this study also aimed to determine whether using the *onebillion* intervention in preschool would have lasting benefits for children when they entered school. Descriptive statistics indicated group differences in CK at follow-up five months later; however, effects were no longer statistically significant, perhaps due in part to low statistical power. Additional research may be beneficial to establish whether changes to the implementation of the intervention in preschool settings might lead to greater maintenance of skills over time.

A further original contribution of this research was to identify whether the *onebillion* tablet-based mathematics intervention may have positive benefits for preschool children's receptive language skills and AtL. Results indicated that there was no significant effect on either measure, although facilitators perceived improvements in specific aspects of AtL for some children, including attention on-task and self-confidence. There were, however, limitations to the sensitivity of both measures in capturing learning gains and potential constraints to the

reliability of teacher-ratings of AtL, suggesting that these findings should be interpreted with caution.

Lastly, this study aimed to illuminate some of the possible mechanisms underpinning the efficacy of the *onebillion* intervention, in order to address “what works, for whom, in what circumstances” (Pawson & Tilley, 1997, p.84), by exploring a range of potential factors that may have affected group outcomes. Analysis of additional embedded data identified that the intervention may not have been universally beneficial for all: children’s prerequisite skills in maths, their attention and their understanding of English may have affected learning gains. Furthermore, a wide variety of other factors were identified that may have affected individual children’s learning experience and may have impacted upon outcomes. The mechanism model proposed in the present research may, therefore, have important implications, as discussed, for the pedagogical use of tablet technology in the early years classroom.

7 References

- Aladé, F., Lauricella, A. R., Beaudoin-Ryan, L., & Wartella, E. (2016). Measuring with Murray: Touchscreen technology and preschoolers' STE< learning. *Computers in Human Behavior*, 62, 433-441.
- Alberto, P.A. & Troutman, A. C. (1986). *Applied Behavior Analysis for Teachers*. (2nd ed.). Columbus, Ohio: Merrill Publishing Company.
- Alexander, R. (2012). Neither national nor a curriculum. *Forum*, 54(3), 369-384.
- American Academy of Pediatrics. (2016). Media and young minds. *Pediatrics*, 138(5), 1-6.
- Anders, Y., Rossbach, H.-G., Weinert, S., Ebert, S., Kuger, S., Lehl, 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.
- Andrews, R. (2005). The place of systematic reviews in education research. *British Journal of Educational Studies*, 53(4), 399-416.
- Ansari, A., & Gershoff, E. (2015). Learning-related social skills as a mediator between teacher instruction and child achievement in Head Start. *Social Development*, 24(4), 699-715.
- Aubrey, C., Godfrey, R., & Dahl, S. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal*, 18(1), 27-46.
- Bailey, D. H., Nguyen, T., Jenkins, J. M., Domina, T., Clements, D. H., & Sarama, J. S. (2016). Fadeout in an early mathematics intervention: Constraining content or preexisting differences? *Developmental Psychology*, 52(9), 1457-1469.
- Baker, S., Gersten, R., & Lee, D.-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *Elementary School Journal*, 103(1), 51-73.

- Ball, J. (2011). *Enhancing Learning of Children from Diverse Language Backgrounds: Mother Tongue-Based Bilingual or Multilingual Education in the Early Years*. Paris, France: UNESCO
- Barbu, O. C., Yaden, D. B., Levine-Donnerstein, D., & Marx, R. W. (2015). Assessing approaches to learning in school readiness. *American Educational Research Association, Open*, 1(3).
- Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, S. C., & Beilock, S. L. (2015). Math at home adds up to achievement in school. *Science*, 350(6257), 196-198.
- Bickford-Smith, A., Wijayatilake, L., & Woods, G. (2005). Evaluating the effectiveness of an early years language intervention. *Educational Psychology in Practice*, 21(3), 161-173.
- Biesta, G. (2010). Pragmatism and the Philosophical Foundations of Mixed Methods Research. In A. Tashakkori, & C. Teddlie, C. (Eds.), *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (2nd ed.) (pp.95-118). London, UK: SAGE.
- Birch, S., Frederickson, N., & Miller, A. (2015). "What do educational psychologists do?" In T. Cline, A. Gulliford, & S. Birch (Eds.). *Educational Psychology* (2nd ed.). Hove, UK: Routledge.
- Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *Elementary School Journal*, 108(2), 115-130.
- Boylan, M., Wolstenholme, C., Maxwell, B., Jay, T., Stevens, A., & Demack, S. (2016). *Longitudinal Evaluation of the Mathematics Teacher Exchange: China-England Interim Research Report*. (DFE-RR559). Sheffield, UK: Sheffield Hallam University.
- Brace, N., Kemp, R., & Snelgar, R. (2012). *SPSS for Psychologists* (5th ed.). London, UK: Palgrave
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.

- Braun, V., & Clarke, V. (2013). *Successful Qualitative Research : A Practical Guide for Beginners*. London, UK: SAGE.
- British Psychological Society. (2009). *Codes of Ethics and Conduct*. Leicester, UK: The British Psychological Society.
- British Psychological Society. (2014). *Codes of Ethics and Conduct*. Leicester, UK: The British Psychological Society.
- Bronsen, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child Behaviour Rating Scale*. Cambridge, Massachusetts: Aht Associates.
- Bruner, J. S. (1966). *Toward a Theory of Instruction*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97-113.
- Bryman, A., & Cramer, D. (2011). *Quantitative Data Analysis with IBM SPSS 17, 18 & 19: A Guide for Social Scientists*. Hove, UK: Routledge.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Carbonneau, K. J., & Marley, S. C. (2015). Instructional guidance and realism of manipulatives influence preschool children's mathematics learning. *Journal of Experimental Education*, 83(4), 495-513.
- Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, 105(2), 380-400.
- Carpenter, H., Papps, I., Bragg, I., Dyson, A., Harris, D., & Kerr, K., Todd, L., & Laing, K. (2013). *Evaluation of Pupil Premium: Research Report*. Manchester: Department for Education.
- Carr (2012). Does Maths achievement h'APP'en when iPads and game-based learning are incorporated into fifth-grade mathematics instruction? *Journal of Information Technology Education*, 11, 269-286.

- Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical Analysis and Practical Implications. *Experimental Psychology*, 56 (4), 236-246.
- Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education*, 105, 14-30.
- Chen, J.-Q., & McNamee, G. D. (2011). Positive approaches to learning in the context of preschool classroom activities. *Early Childhood Education Journal*, 39(1), 71-78.
- Chen, L., Bae, S. R., Battista, C., Qin, S., Chen, T., Evans, T. M., & Menon, V. (2018). Positive attitude toward math supports early academic success: Behavioral evidence and neurocognitive mechanisms. *Psychological Science*, 29(3), 390-402.
- Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88-113.
- Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283-292.
- Chow, J. C., & Ekholm, E. (in press). Language domains differentially predict mathematics performance in young children. *Early Childhood Research Quarterly*.
- Clarke, L., & Abbot, L. (2016). Young pupils', their teacher's and classroom assistants' experiences of iPads in a Northern Ireland school: "Four and five years old, who would have thought they could do that?" *British Journal of Educational Technology*, 47(6), 1051-1064.
- Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, 12(3), 297-298.
- Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Cary, M. S. (2016). Examining the efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Learning Disabilities*, 49(2), 152-165.

- Clark-Carter, D. (2007). Effect size and statistical power in psychological research, *The Irish Journal of Psychology*, 28(1-2), 3-12.
- Clements, D. H. & Sarama, J. (2009). *Learning and Teaching Early Math: The Learning Trajectories Approach*. New York: Routledge.
- Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal*, 50(4), 812-850.
- Cohen, (1988). *Statistical Power for the Behavioural Sciences* (2nd ed.). New York: Academic Press
- Cook, J. A. (2009). The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. *Trials*, 10, 1-9.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and Analysis for Field Settings*. Chicago: Rand McNally
- Cook, B. G., Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children*, 79 (2), 135-144.
- Coolican, H. (2014). *Research Methods and Statistics in Psychology* (6th ed.). London : Routledge.
- Cordes, C., & Miller, E. (2000). *Fool's gold: A critical look at computers in childhood*. Alliance for Childhood. Retrieved 27 07 2017, from http://drupal6.allianceforchildhood.org/fools_gold
- 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), 75-98.
- Crawford, C., & Cribb, J. (2013). *Reading and Maths Skills at Age 10 and Earnings in Later Life: A Brief Analysis using the British Cohort Study*. Retrieved 26 07 2017, from <https://www.ifs.org.uk/publications/6624>.
- Creswell, J. W. (2014). *Research Design : Qualitative, Quantitative, and Mixed Method Approaches* (4th ed.). London : SAGE.

Deci, E. L., & Ryan, R. M. (2000). The " what" and " why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268.

Department for Business, Innovation & Skills. (2012). *The 2011 Skills for Life Survey: A Survey of Literacy, Numeracy and ICT Levels in England*. Retrieved 26 07 2017, from <https://www.gov.uk/government/publications/2011-skills-for-life-survey>.

Department for Communities & Local Government. (2015). *The English Indices of Deprivation 2015: Statistical Release*. Retrieved 18 05 2018, from <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>

Department for Education. (2017a). *Early Years Foundation Stage Profile Results: 2016-2017*. Retrieved 19 05 2018, from <https://www.gov.uk/government/statistics/early-years-foundation-stage-profile-results-2016-to-2017>

Department for Education (2017b). *Statutory framework for the early years*. Retrieved 26 07 2017, from <https://www.gov.uk/government/publications/early-years-foundation-stage-framework--2>

Department for Education. (2016). *Schools, Pupils and their Characteristics: January 2017*. Retrieved 19 05 2018, from <https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2017>

Department for Education. (2013). *The National Curriculum in England: Key Stages 1 and 2 Framework Document*. (DFE-00178-2013). London: Department for Education Publications.

DiPerna, J. C., Lei, P.-W., & Reid, E. E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the Early Childhood Longitudinal Study--Kindergarten Cohort. *Journal of Educational Psychology*, 99(2), 369-379.

- Doabler, C. T., & Fien, H. (2013). Explicit mathematics instruction: What teachers can do for teaching students with mathematics difficulties? *Intervention in School and Clinic, 48*(5), 276-285.
- Doabler, C. T., Fien, H., Nelson-Walker, N. J., & Baker, S. K. (2012). Evaluating three elementary mathematics programs for presence of eight research-based instructional design principles. *Learning Disability Quarterly, 35*(4), 200-211.
- Dobbs, J., Doctoroff, G. L., Fisher, P. H., & Arnold, D. H. (2006). The association between preschool children's socio- emotional functioning and their mathematical skills. *Journal of Applied Developmental Psychology, 27*(2), 97-108.
- Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 324-332.
- Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Developmental Science, 11*(5), 650-654.
- Dowker, A. (2009). *What Works for Children with Mathematical Difficulties? The effectiveness of intervention schemes*. Nottingham, UK: Department for Children Schools and Families Publications.
- Dugard, P., & Todman, J. (1995). Analysis of pre-test and post-test control group designs in educational research. *Educational Psychology, 15*(2), 181-198.
- Dulaney, A., Vasilyeva, M., & O'Dwyer, L. (2015). Individual differences in cognitive resources and elementary school mathematics achievement: Examining the roles of storage and attention. *Learning and Individual Differences, 37*, 55-63.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446.
- Education Endowment Foundation (2015). *Analysis Guidance*. Retrieved 18 05 2018, from <https://educationendowmentfoundation.org.uk/projects-and->

evaluation/evaluating-projects/evaluator-resources/setting-up-an-evaluation/.

Elliot, C. D., & Smith, P. (2011). *The British Ability Scales 3*. London, UK: GL Assessment.

Falloon, G. (2016). An analysis of young students' thinking when completing basic coding tasks using Scratch Jnr. On the iPad. *Journal of Computer Assisted Learning*, 32(6), 576-593.

Fawcett, L. M., & Garton, A. F. (2005). The effect of peer collaboration on children's problem-solving ability. *British Journal of Educational Psychology*, 75(2), 157-169.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307-314.

Feilzer, Y. M. (2010). Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm. *Journal of Mixed Methods Research*, 4(1), 6-16.

Fien, H., Doabler, C. T., Nelson, N. J., Kosty, D. B., Clarke, B., & Baker, S. K. (2016). An examination of the promise of the NumberShire Level 1 gaming intervention for improving student mathematics outcomes. *Journal of Research on Educational Effectiveness*, 9(4), 635-661.

Fox, M. (2011). Practice-based evidence - Overcoming insecure attachments. *Educational Psychology in Practice*, 27(4), 325-335.

Fuchs, L. S., Fuchs, D., Compton, D. L. (2011). The early prevention of mathematics difficulty. *Journal of Learning Disabilities*, 45(3), 257-269.

Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *Elementary School Journal*. 101(5), 495-510.

Gagné, R. M. (1968). Contributions of learning to human development. *Psychological Review*, 75(3), 177-191.

- Gelman, R. (1972). Logical capacity of very young children: Number invariance rules. *Child Development*, 43(1), 75-90.
- Gelman, R. (2000). The epigenesis of mathematical thinking. *Journal of Applied Developmental Psychology*, 21(1), 27-37.
- Gersten, R. (2016). What we are learning about mathematics interventions and conducting research on mathematics interventions. *Journal of Research on Educational Effectiveness*, 9(4), 684-688.
- Gersten, R., & Carnine, D. (1984). Direct instruction mathematics: A longitudinal evaluation of low-income elementary school students. *The Elementary School Journal*, 84(4), 395-407.
- Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, 79(3), 1202-1242.
- Gifford, S., Black, J., & Griffiths, R. (2015). Making numbers: Where are we now. *Proceedings of the British Society for Research into Learning Mathematics*, 35(2), 19-24.
- Gifford, S., & Rockliffe, F. (2012). Mathematics difficulties: Does one approach fit all? *Research in Mathematics Education*, 14(1), 1-15.
- Ginsburg, H. P., & Baroody, A. J. (2003). *Test of Early Mathematics Ability, Third Edition (TEMA-3)*. Austin, TX: Pro-Ed.
- Ginsburg, H. P., & Amit, M. (2008). What is teaching mathematics to young children? A theoretical perspective and case study. *Journal of Applied Developmental Psychology*, 29(4), 274-285.
- Ginsburg, H. P., & Golbeck, S. L. (2004). Thoughts on the future of research on mathematics and science learning and education. *Early Childhood Research Quarterly*, 19(1), 190-200.
- Gough, D. (2007). Weight of Evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2), 213-228.

- Gough, D., Oliver, S., & Thomas, J. (2012). *An Introduction to Systematic Reviews*. London : SAGE.
- Gough, D., Thomas, J., & Oliver, S. (2012). Clarifying differences between review designs and methods. *Systematic Reviews*, 1(1), 28.
- Greany, T., Barnes, I., Mostafa, T., Pensiero, N., & Swensson, C. (2016). *Trends in Maths and Science Study (TIMSS): National Report for England*. Retrieved 26 07 2017, from <https://www.gov.uk/government/publications/timss-2015-national-report-for-england>
- Green, J., Roberts, H., Petticrew, M., Steinbach, R., Goodman, A., Jones, A., & Edwards, P. (2015). Integrating quasi-experimental and inductive designs in evaluation: A case study of the impact of free bus travel on public health. *Evaluation*, 21(4), 391-406.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: SAGE.
- Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: Relations between parents' input and children's cardinal-number knowledge. *Developmental Science*, 14(5), 1021-1032.
- Haake, M., Husain, L., & Gulz, A. (2015). Supporting early math: Rationales and requirements for high quality software. *Journal of Computers in Mathematics and Science Teaching*, 34(4), 409-429.
- Hachey, A. C. (2013). Early childhood mathematics education revolution. *Early Education and Development*, 24(4), 419-430.
- Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The Fourth R-Research in the Classroom*. Ohio: Charles. E. Merrill.
- Haßler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: a critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning*, 32(2), 139-156.

- Health & Care Professions Council. (2012). *Standards of Proficiency: Practitioner Psychologists*. London: Health and Care Professions Council
- Hemmings, B., Grootenboer, P., & Kay, R. (2011). Predicting mathematics achievement: The influence of prior achievement and attitudes. *International Journal of Science and Mathematics Education*, 9(3), 691-705.
- Henderson, S., & Yeow, J. (2012). iPad in education: A case study of iPad adoption and use in a primary school. In *Proceedings of the 2012 45th Hawaii International Conference on System Sciences* (pp.78-87). Washington, DC: IEEE Computer Society.
- 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.
- Hieftje, K., Pendergrass, T., Kyriakides, T. C., Gilliam, W., & Fiellin, L. (2017). An evaluation of an educational video game on mathematics achievement in first grade students. *Technologies*, 5(2), 1-8.
- Higgins, S., Kokotsaki, D., & Coe, R. (2012). *The Teaching and Learning Toolkit: Technical Appendices*. Retrieved 15 03 2016, from <https://educationendowmentfoundation.org.uk>.
- Higgins, S., Xiao, Z., & Katsipatakis, M. (2012). *The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation*. London, UK: EEF.
- Hindman, A. H., Erhart, A. C., & Wasik, B. A. (2012). Reducing the Matthew Effect: Lessons from the "ExCELL" Head Start Intervention. *Early Education and Development*, 23(5), 781-806.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brookes Publishing.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26, 55-88.

- 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.
- Horner, R., Sugai, G., & Fixsen, D. (2017). Implementing effective educational practices at scales of social importance. *Clinical Child and Family Psychology Review*, 20(1), 25-35.
- Howell, S., & Kemp, C. (2006). An international perspective of early number sense: Identifying components predictive of difficulties in early mathematics achievement. *Australian Journal of Learning Disabilities*, 11(4), 197-207.
- Howell, S. C., & Kemp, C. R. (2010). Assessing preschool number sense: Skills demonstrated by children prior to school entry. *Educational Psychology*, 30(4), 411-429.
- Hubber, P. J., 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, 1-4.
- Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General*, 123(3), 284-296.
- Joffe, H. (2012). Thematic Analysis. In D. Harper, & A. Thompspon (Eds.). *Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners* (pp.209-223). Chichester, UK: Wiley-Blackwell.
- Johnson, C., Hart, R., & Colwell, J. (2014). *Steps to Engaging Young Children in Research. Volume 1: The Guide*. Retrieved 19 05 2018, from <http://www.comminit.com/early-child/content/steps-engaging-young-children-research-volume-1-guide>

- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- Jordan, N. C., Glutting, J., Dyson, N. Hassinger-Das, B., & Casey, I. (2012). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology*, 104(3), 647-660.
- 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.
- Kagan, S. L., Moore, E., & Bredakamp, S. (Eds.). (1995). *Reconsidering Children's Early Development and Learning: Toward Common Views and Vocabulary*. Washington, DC: National Education Goals Panel.
- Kim, P., Buckner, E., Kim, H., Makany, T., Taleja, N., & Parikh, V. (2012). A comparative analysis of a game-based mobile learning model in low-socioeconomic communities of India. *International Journal of Educational Development*, 32(2), 329-340.
- 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.
- Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher math talk. *Developmental Psychology*, 42(1), 59-69.

- Kosko, K., & Ferdig, R. (2016). Effects of a tablet-based mathematics application for pre-school children. *Journal of Computers in Mathematics and Science Teaching*, 35(1), 61-79.
- Krinzinger, H., Kaufmann, L., & Willmes, K. (2009). Math anxiety and math ability in early primary school years. *Journal of Psychoeducational Assessment*, 27(3), 206-225.
- Kucirkova, N. (2014). iPads in early education: Separating assumptions and evidence. *Frontiers in Psychology*, 5(715), 1-3.
- Larkin, K. (2013). Mathematics education: Is there an app for that? In V. Steinle, L. Ball, & C. Bordini (Eds.), *Mathematics education: Yesterday, today, and tomorrow* (pp. 426–433). Melbourne, Australia: MERGA.
- LeBuffe, P.A., & Naglieri, J. A. with the Devereux Center for Resilient Children (2012). *The Devereux Early Childhood Assessment for Preschoolers - Second Edition*. Lewisville, NC: Kaplan Early Learning Company.
- LeFevre, J. A., Fast, L., Skwarchuk, S. I., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, 81(6), 1753-1767.
- Lepola, J., Lynch, J., Kiuru, N., Laakkonen, E., & Niemi, P. (2016). Early oral language comprehension, task orientation, and foundational reading skills as predictors of Grade 3 reading comprehension. *Reading Research Quarterly*, 51(4), 373-390.
- Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge? *Developmental Psychology*, 46(5), 1309-1319.
- Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school student's mathematics learning. *Educational Psychology Review*, 22(3), 215-243.

- Li-Grining, C. P., Votruba-Drzal, E., Maldonado-Carreno, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology*, 46(5), 1062-1077.
- Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: Implications for the early years and beyond. *International Journal of Language & Communication Disorders*, 37(1), 3-15.
- Lovato, S. B., & Waxman, S. R. (2016). Young children learning from touch screens: Taking a wider view. *Frontiers in psychology*, 7, 1078.
- Magliaro, S., Lockee, B., & Burton, J. (2005). Direct instruction revisited: A key model for instructional technology. *Educational Technology Research and Development*, 53(4), 41-55.
- Main, S., O'Rourke, J., Morris, J., & Dunjey, H. (2016). Focus on the journey, not the destination: Digital games and students with disability. *Issues in Educational Research*, 26(2), 315-331.
- Marcelino, L., de Sousa, Ó., Cruz, V., & Lopes, A. (2012). Multi-year longitudinal investigation of children's early mathematics development. *Procedia - Social and Behavioral Sciences*, 69, 1911-1920.
- Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J.C., Lahmar, J., Scott, F., Davenport, A., Davis, S., French, K., Piras, M., Thornhill, S., Robinson, P. and Winter, P. (2015) *Exploring Play and Creativity in Pre-Schoolers' Use of Apps: Final Project Report*. Retrieved 27 07 17, from http://www.techandplay.org/reports/TAP_Final_Report.pdf.
- Martens, B. K., & Witt, J. C. (2004). Competence, persistence and success: The positive psychology of behavioral skill instruction. *Psychology in the Schools*, 4(1), 19-30.
- Matthews, J., & Seow, P. (2007). Electronic paint: Understanding children's representation through their interactions with digital paint. *International Journal of Art & Design Education*, 26(3), 251-263.
- Mattoon, C., Bates, A., Shifflet, R., Latham, N., & Ennis, S. (2015). Examining computational skills in prekindergarteners: The effects of traditional and

digital manipulatives in a prekindergarten classroom. *Early Childhood Research & Practice*, 17(1).

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59(1), 14-19.

Mazzocco, M. M. M. (2009). An introduction to the special issue: Pathways to mathematical learning difficulties and disabilities. *Developmental Disabilities Research Reviews*, 15(1), 1-3.

McCall, G. J. (1984). Systematic field observation. *Annual Review of Sociology*, 10, 263-282.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, 21(4), 471-490.

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: The role of learning-related social skills. *Early Childhood Research Quarterly*, 15(3), 307-329.

McDermott, P. A., Fantuzzo, J. W., Warley, H. P., Waterman, C., Angelo, L. E., Gadsden, V. L., & Sekino, Y. (2011). Multidimensionality of teachers' graded responses for preschoolers' stylistic learning behavior: The learning-to-learn scales. *Educational and Psychological Measurement*, 71(1), 148-169.

McDermott, P. A., Green, L. F., Francis, J. M., & Stott, D. H. (2000). *Preschool Learning Behaviors Scale*. Philadelphia: Edumetric and Clinical Science.

McGarrigle, J., & Donaldson, M. (1974). Conservation accidents. *Cognition*, 3(4), 341-350.

Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2008). Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school. *Journal of Social Issues*, 64(1), 95-114.

- Merewether, J., & Fleet, A. (2014). Seeking children's perspectives: A respectful layered research approach. *Early Child Development and Care*, 184(6), 897-914.
- Merrett, F., & Wheldall, K. (1986). Observing pupils and teachers in classrooms (OPTIC): A behavioural observation schedule for use in schools. *Educational Psychology*, 6(1), 55-70.
- Mertens, D. M. (2015). *Research and Evaluation in Education and Psychology : Integrating Diversity with Quantitative, Qualitative, and Mixed Methods* (4 ed.). London : SAGE.
- Moher, D., Liberati, A., Tetzlaff, J., Altman., D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264-270.
- Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), 48-76.
- Morgan, D. L. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry*, 20(8), 1045-1053.
- Moyer-Packenham, P. S., Shumway, J. F., Bullock, E., Tucker, S. I., Anderson-Pence, K. L., Westenskow, A., . . . Jordan, K. (2015). Young children's learning performance and efficiency when using virtual manipulative mathematics iPad apps. *Journal of Computers in Mathematics and Science Teaching*, 34(1), 41-69.
- National Centre for Excellence in the Teaching of Mathematics. (2014). *Mastery Approaches to Mathematics and the new National Curriculum*. Sheffield: NCETM.
- Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Childhood Research Quarterly*, 36(3), 550-560.

- Napoli, A. R., & Purpura, D. J. (2018). The home literacy and numeracy environment in preschool: cross-domain relations of parent-child practices and child outcomes. *Journal of Experimental Child Psychology*, 166, 581-603.
- Nesdale, D., & Flessner, D. (2001). Social identity and the development of children's group attitudes. *Child Development*, 72(2), 506-517.
- Neumann, M. M., & Neumann, D. L. (2014). Touch screen tablets and emergent literacy. *Early Childhood Education Journal*, 42(4), 231-239.
- Nordstrum, L. E., LeMahieu, P. G., & Berrena, E. (2017). Implementation science: Understanding and finding solutions to variation in program implementation. *Quality Assurance in Education: An International Perspective*, 25(1), 58-73.
- Norman, G. (2010). Likert Scales, Levels of Measurement and the "Laws" of Statistics. *Advances in Health Sciences Education*, 15(5), 625-632.
- Nuzzo, R. L. (2016). Statistical Power. *PM&R*, 8(9), 907-912.
- onebillion. (2016a). Maths Age 3-5 (Version 1.4) [Mobile Application Software] Retrieved 24 07 17, from <https://onebillion.org/apps>.
- onebillion. (2016b). Maths Age 4-6 (Version 1.2) [Mobile Application Software] Retrieved 24 07 17, from <https://onebillion.org/apps>.
- onebillion. (2018). *Apps to Help Your Child Reach their Full Potential in Maths*. Retrieved 24 07 17, from <https://onebillion.org/apps>.
- Ofcom. (2016). *Children and Parents: Media Use and Attitudes Report*. Retrieved 27 07 17, from <https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-parents-nov16>.
- Ofsted. (2017). *Bold Beginnings: The Reception Curriculum in a Sample of Good and Outstanding Primary Schools*. Retrieved 18 05 2018, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/663560/28933_Ofsted_-_Early_Years_Curriculum_Report_-_Accessible.pdf

Ofsted. (2012). *Mathematics: Made to Measure*. Crown Copyright: OFSTED Publications Centre.

Organisation for Economic Co-Operation and Development. (2017). *Starting Strong 2017: Key OECD Indicators on Early Childhood Education and Care*. Retrieved 19 05 2018, from <http://www.oecd.org/education/starting-strong-2017-9789264276116-en.htm>

Organisation for Economic Co-Operation and Development. (2016). *PISA 2015 Results in Focus*. Retrieved 26 07 2017, from <https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf>

Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J (in press). Maths app counts for school success. *Journal of Educational Psychology*.

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., Neves, M., Gulliford, A., & Pitchford, N. J. (In preparation). Supporting early maths achievement for bilingual children with interactive maths apps.

Oxford University Press (2018). *Numicon: Making Maths Real*. Retrieved 18 08 18, from <https://global.oup.com/education/content/primary/series/numicon/>

Pallant, J. (2016). *SPSS Survival Manual: A Step by Step Guide to Data Analysis using IBM SPSS*. Maidenhead : McGraw-Hill Education.

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.

- Park, J., & Lee, J. (2015). Dyadic collaboration among preschool-age children and the benefits of working with a more socially advanced peer. *Early Education and Development*, 26(4), 574-593.
- Parsons, S., & Bynner, J. (2006). *Does Numeracy Matter More?* London, UK: National Research and Development Centre.
- Pavio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford, UK: Oxford University Press.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London, UK: SAGE.
- Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford, UK: Blackwell.
- Piaget, J. P. (1969). *The Child's Conception of Number*. Norwich, UK: Fletcher & Son. (Original work published 1941)
- Pitchford, N. J. (2015). Development of early mathematical skills with a tablet intervention: A randomized control trial in Malawi. *Frontiers in Psychology*, 6, 12.
- Pitchford, N. J., Kamchedzera, E., Hubber, P. J., & Chigeda, A. L. (2018). Interactive apps promote learning of basic mathematics in children with special educational needs and disabilities. *Frontiers in Psychology*, 9, 1-14.
- Pitchford, N. J., & Outhwaite, L. A. (In preparation). Impact of screen time and screen content on child development.
- Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology*, 52(2), 126-136.
- Powell, S. R., & Driver, M. K. (2015). The influence of mathematics vocabulary instruction embedded within addition tutoring for first-grade students with mathematics difficulty. *Learning Disability Quarterly*, 38(4), 221-233.
- Praet, M., & Desoete, A. (2014). Enhancing young children's arithmetic skills through non-intensive, computerised kindergarten interventions: A

- randomised controlled study. *Teaching and Teacher Education*, 39, 56-65.
- Praet, M., Titeca, D., Ceulemans, A., & Desoete, A. (2013). Language in the prediction of arithmetics in Kindergarten and Grade 1. *Learning and Individual Differences*, 27, 90-96.
- Prieto, L., Arreguin-Anderson, M. G., Yuen, T. T., Ek, L. D., Sanchez, P., Machado-Casas, M., & Garcia, A. (2016). Four cases of a sociocultural approach to mobile learning in La Clase Magica, an afterschool technology club. *Interactive Learning Environments*, 24(2), 345-356.
- Purpura, D. J., & Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology*, 122, 104-121.
- Purpura, D. J., & Reid, E. E. (2016). Mathematics and language: Individual and group differences in mathematical language skills in young children. *Early Childhood Research Quarterly*, 36, 259-268.
- Razza, R., Martin, A., & Brooks-Gunn, J. (2015). Are approaches to learning in kindergarten associated with academic and social competence similarly? *Journal of Research and Practice in Children's Services*, 44(6), 757-776.
- Rittle-Johnson, B., Fyfe, E. R., Hofer, K. G., & Farran, D. C. (2017). Early math trajectories: Low-income children's mathematics knowledge from ages 4 to 11. *Child Development*, 88(5), 1727-1742.
- Roberts, M. L., Marshall, J., Nelson, J. R., & Albers, C. A. (2001). Curriculum-based assessment procedures embedded within functional behavioral assessments: Identifying escape-motivated behaviors in a general education classroom. *School Psychology Review*, 30(2), 264-277.
- Robson, C. (2011). *Real World Research : A Resource for Users of Social Research Methods in Applied Settings* (3rd ed.). Chichester : Wiley.
- Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the Worker*. Cambridge, MA: Harvard University.

- Roschelle, J., Rafanan, K., Estrella, G., Nussbaum, M., & Claro, S. (2010). From handheld collaborative tool to effective classroom module: Embedding CSCL in a broader design framework. *Computers & Education, 55*(3), 1018-1026.
- Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research, 95*(5), 259-272.
- Roulstone, S., Law, J., Rush, R., Clegg, J., & Peters, T. (2011). *Investigating the Role of Language in Children's Early Educational Outcomes*. Department for Education Research Report 134.
- Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics, 2*(3), 808-840.
- Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers* (3rd ed.). London : SAGE.
- Salminen, J. B., Koponen, T. K., Leskinen, M., Poikkeus, A.-M., Aro, M. T. (2015). Individual variance in responsiveness to early computerised mathematics intervention. *Learning and Individual Differences, 43*, 124-131.
- Sarama, J., Lange, A. A., Clements, D. H., Wolfe, C. B. The impacts of an early mathematics curriculum on oral language and literacy. *Early Childhood Research Quarterly, 27*(3), 489-502.
- 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. (2016). Math Shelf: A randomized trial of a prekindergarten tablet

number sense curriculum. *Early Education and Development*, 27(1), 74-88.

Schrank, F.A., McGrew, K. S., & Mather, N. (2015). *Woodcock-Johnson IV Tests of Early Cognitive and Academic Development*. Rolling Meadows, IL: Riverside.

Schunk, D. H. (1983). Ability versus effort attributional feedback: Differential effects on self-efficacy and achievement. *Journal of Educational Psychology*, 75(6), 848-856.

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology*, 19, 107-122.

Shanley, L., Cary, M. S., Clarke, B., Guerreiro, M. A., & Thier, M. (2017). Instructors' technology experience and iPad delivered intervention implementation: a mixed methods replication study. *Educational Technology Research and Development*, 65(4), 815-830.

Shapiro, L. R., & Solity, J. (2008). Delivering phonological and phonics training within whole-class teaching. *British Journal of Educational Psychology*, 78(4), 597-620.

Sigman, A. (2012). Time for a view on screen time. *Archives of Disease in Childhood*, 97(11), 935.

Sims, D. M., Purpura, D. J., & Lonigan, C. J. (2016). The relation between inattentive and hyperactive/impulsive behaviours and early mathematics skills. *Journal of Attention Disorders*, 20(8), 704-714.

Skinner, B. F. (1953). *Science and Human Behaviour*. New York, NY: Free Press

Skwarchuk, S.-L., Sowinski, C., & Lefevre, J.-A. (2014). Formal and informal home learning activities in relation to children's early numeracy and literacy skills: The development of a home numeracy model. *Journal of Experimental Child Psychology*, 121, 63-84.

- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15-21.
- Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research*, 78(3), 427-515.
- Son, L. K., & Simon, D. K. (2012). Distributed learning: Data, metacognitions and educational implications. *Educational Psychology Review*, 24, 379-399.
- 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, 12.
- Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, 19(1), 99-120.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-406.
- Stubbé, H., Badri, A., Telford, R., van der Hulst, A., & van Joolingen, W. (2016). E-learning Sudan, formal learning for out-of-school children. *Electronic Journal of E-Learning*, 14(2), 136-149.
- Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from likert-type scales. *Journal of Graduate Medical Education*, 5(4), 541.
- Susperreguy, M. I., & Davis-Kean, P. E. (2016). Maternal math talk in the home and math skills in preschool children. *Early Education and Development*, 1-17.
- Szkudlarek, E., & Brannon, E. M. (2017). Does the approximate number system serve as a foundation for symbolic mathematics? *Language Learning and Development*, 13(2), 171-190.

- Taggart, B., Sylva, K., Melhuish, E., Sammons, P., & Siraj, I. (2015). *Effective Pre-school, Primary and Secondary Education Project (EPPSE 3-16+): How Pre-school Influences Children and Young People's Attainment and Developmental Outcomes over Time. Research Brief*. Retrieved 18 05 2018, from http://dera.ioe.ac.uk/23344/1/RB455_Effective_pre-school_primary_and_secondary_education_project.pdf
- Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. London, UK: Academic Press for European Association of Experimental Social Psychology.
- Tajfel, H., & Turner, J. (1979) An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks/Cole.
- Teddlie, C. & Tashakkori, A. (2009). *Foundations of Mixed Methods Research : Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. London : SAGE.
- Teepe, R. C., Molenaar, I., & Verhoeven, L. (2016). Technology-enhanced storytelling stimulating parent-child interaction and preschool children's vocabulary knowledge. *Journal of Computer Assisted Learning*, 33(2), 123-136.
- Thompson, I. (2012). The draft national curriculum for primary mathematics. *Mathematics Teaching*, 230, 15-16.
- Threlfall, J., & Bruce, B. (2005). "Just" counting: Young children's oral counting and enumeration. *European Early Childhood Education Research Journal*, 13(2), 63-77.
- Toll, S. W. M., & Van Luit, J. E. H. (2014). The developmental relationship between language and low early numeracy skills throughout kindergarten. *Exceptional Children*, 81(1), 64-78.
- Torgerson, C., & Torgerson, D. (2003). The design and conduct of randomised controlled trials in education: Lessons from health care. *Oxford Review of Education*, 29(1), 67-80.

University of Nottingham. (2015). *Code of Research Conduct and Research Ethics*. Nottingham: University of Nottingham.

Valle-Lisboa, J., Cabana, Á., Eisinger, R., Mailhos, Á., Luzardo, M., Halberda, J., & Maiche, A. (2016). Cognitive abilities that mediate SES's effect on elementary mathematics learning: The Uruguayan tablet-based intervention. *Prospects: Quarterly Review of Comparative Education*, 46(2), 301-315.

van Breukelen, G. J. P. (2013). ANCOVA Versus CHANGE From Baseline in nonrandomized Studies: The Difference. *Multivariate Behavioral Research*, 48(6), 895-922.

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.

Verenikina, I., J. Herrington, R. Peterson & Mantei, J. (2010). Computers and play in early childhood: Affordances and limitations. *Journal of Interactive Learning Research*, 21(1), 19-159.

Vukovic, R. K., & Lesaux, N. K. (2013). The language of mathematics: Investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology*, 115(2), 227-244.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press

Walter-laager, C., Brandenburg, K., Tinguely, L., Pfiffner, M. R., & Moschner, B. (2016). Media-assisted language learning for young children: Effects of a word-learning app on the vocabulary acquisition of two-year-olds. *British Journal of Educational Technology*, 48(4), 1062-1072.

Wang, A. H., Firmender, J. M., Power, J. R., & Byrnes, J. P. (2016). Understanding the program effectiveness of early mathematics interventions for prekindergarten and kindergarten environments: A meta-analytic review. *Early Education and Development*, 1-22.

- Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue. *Educational Researcher*, 43(7), 352-360.
- Weiss, R. S. (1994). *Learning from Strangers : The Art and Method of Qualitative Interview Studies*. New York : Free Press.
- Wheater, R., Durbin, B., McNamara, S., & Classick, R. (2016). *Is Mathematics Education in England Working for Everyone? NFER Analysis of the PISA Performance of Disadvantaged Pupils*. Slough, NFER.
- Wigg, E. H., Secord, W., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals – Preschool 2 UK*. Oxford, UK: Pearson Clinical.
- Williford, A. P., Maier, M. F., Downer, J. T., Pianta, R. C., & Howes, C. (2013). Understanding how children's engagement and teachers' interactions combine to predict school readiness. *Journal of Applied Developmental Psychology*, 34(6), 299-309.
- Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55(1), 235-269.
- Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Child Psychiatry*, 17, 89-100.
- Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, 89(1), B15-B25.
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74(1), B1-B11.
- Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, 8(1), 88-101.
- Yee, M. D., & Brown, R. (1992). Self-evaluations and intergroup attitudes in children aged three to nine. *Child Development*, 63(3), 619-629.
- Yelland, N., & Masters, J. (2007). Rethinking scaffolding in the information age. *Computers and Education*, 48(3), 362-382.
- Zimmerman, I. L., Pond, R. E., & Steiner, V. G. (2011). *Preschool Language Scale – Fifth Edition*. Oxford, UK: Pearson Clinical.

Zur, O., & Gelman, R. (2004). Young children can add and subtract by predicting and checking. *Early Childhood Research Quarterly*, 19(1), 121-137.

8 Appendices

8.1 Systematic Literature Review: Reasons for excluding articles following title and abstract screening

Reasons for exclusion of articles based upon titles and abstracts are displayed below, with reference to the eligibility criteria in Section 2.6.2.

Reason for exclusion	Number of Articles
Article was not an evaluation of a mathematics-specific tablet-based intervention (e.g. a review paper, or a study focusing on an unrelated topic)	91
Intervention was not a clearly defined supplemental CAI programme (e.g. a number of different intervention programmes with different instructional features or used technology as a tool to support the wider lesson)	4
All participants were over the age of 6 years	22
Study did not assess quantitative outcomes on early mathematical understanding and skills	3
Participants were restricted to children with specific diagnoses of additional needs or recognised disabilities (e.g. children with hearing impairment)	2
Study did not include a pre- and post-test design, with a control or comparison group (not accessing a clearly defined tablet-intervention programme but other technology-based and/or maths activities)	10
Total	132

8.2 Systematic Literature Review: Excluded full-text articles

The full texts articles that were excluded from the SLR are displayed below, alongside the reasons for their exclusion.

Author	Reason for exclusion
Aladé, Lauricella, Beaudoin-Ryan and Wartella (2016)	Study conducted during a single intervention session
Falloon (2016)	Did not measure mathematical outcomes
Kim et al. (2012)	The intervention did not use tablet technology
Mattoon et al. (2015)	The intervention involved multiple programmes with different instructional features
Main, O'Rourke and Morris (2016)	The intervention did not use tablet technology
Pitchford et al. (2018)	No control group included within the design
Prieto et al. (2016)	Did not evaluate a mathematics intervention
Roschelle, Rafanan, Estrella, Nussbaum and Claro ((2010)	Participants over 6 years
Stacy, Cartwright, Arwood, Canfield and Kloos (2017)	Did not measure mathematical outcomes
Stubbé, Badri, Telford, van der Hulst and van Joolingen (2016)	Participants over 6 years
Valle-Lisboa et al. (2016)	No control group included within the design

In addition, two studies from one record included in the review (Outhwaite et al., 2017), referred to as Study 2 and 3 by the authors, were excluded from the SLR as they did not include a control group.

8.3 Weight of Evidence Criteria Judgements for Systematic Literature Review

Weight of Evidence Judgement A: Generic judgement about the quality of the research design

The number of criteria met were totalled and these scores were used to assign categorical labels according to the following classification system: **High** (5-6), **Medium** (3-4), **Low** (0-2)

Criteria used for this review	Berkowitz et al. (2015)	Hieftje et al. (2017)	Kosko and Ferdig (2016)	Outhwaite et al. (2017) Study 1	Outhwaite et al. (2017) Study 4	Outhwaite et al. (in in pres)	Park et al. (2016)	Pitchford (2015)	Schacter and Jo (2016)	Schacter et al. (2016)	Schacter and Jo (2017)	van der Ven et al. (2017)
Sample clearly specified (inc. setting/child characteristics)	N	N	N	Y	Y	Y	N	Y	Y	Y	N	N
Clear specification of the intervention (inc. number/length sessions, duration, software content)	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Intervention fidelity reported (e.g. number of sessions individual children completed)	Y	Y	N	N	N	N	N	N	N	N	N	N
Standardised measures	Y	Y	N	N	Y	Y	Y	N	N	N	N	N
Quantitative data presented clearly	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Limitations acknowledged	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Rating	M	H	L	M	H	H	M	M	M	M	M	M

Weight of Evidence Judgement B: Specific judgement of the appropriateness of method to answer the review question

The number of criteria met were totalled and these scores were used to assign categorical labels according to the following classification system: **High** (5-6), **Medium** (3-4), **Low** (0-2)

Criteria used for this review	Berkowitz et al. (2017)	Hiefje et al. (2017)	Kosko and Ferdig (2016)	Outhwaite et al. (2017) Study 1	Outhwaite et al. (2017) Study 4	Outhwaite et al. (in press)	Park et al. (2016)	Pitchford (2015)	Schacter and Jo (2016)	Schacter et al. (2016)	Schacter and Jo (2017)	van der Ven et al. (2017)
RCT or Cluster RCT	Y	Y	Y	N	N	Y	Y	Y	N	Y	Y	Y
Treatment as usual or other mathematics programme control group	N	N	Y	Y	Y	Y	N	Y	Y	N	N	Y
Comparison group with access to technology	Y	Y	N	N	N	N	Y	Y	N	Y	Y	N
Significance testing	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Control and intervention groups shown to be equivalent on baseline mathematics (no sig. differences) or controlled for differences between groups in statistical analysis	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y
Longitudinal follow up included after at least 1 month	N	N	N	Y	N	N	N	N	N	N	N	Y
Rating	M	M	M	M	L	M	M	H	M	M	M	H

Weight of Evidence Judgement C: *Relevance of the study in answering the review question*

The number of criteria met were totalled and these scores were used to assign categorical labels according to the following classification system: **High** (4-5), **Medium** (2-3), **Low** (0-1)

Criteria used for this review	Berkowitz et al. (2015)	Hieftje et al. (2017)	Kosko and Ferdig (2016)	Outhwaite et al. (2017) Study 1	Outhwaite et al. (2017) Study 4	Outhwaite et al. (in press)	Park et al. (2016)	Pitchford (2015)	Schacter and Jo (2016)	Schacter et al. (2016)	Schacter and Jo (2017)	van der Ven et al. (2017)
Includes <u>only</u> children aged 6 years and below	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	N
Large sample size ≥ 100 at post-test	Y	Y	N	N	N	Y	Y	Y	Y	N	Y	N
Conducted in an educational setting and during the school day	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Teaching staff facilitator	N	N/S	N	Y	Y	Y	N/S	Y	Y	N	Y	Y
Conducted in the UK	N	N	N	Y	Y	Y	N	N	N	N	N	N
Rating	L	L	L	M	H	H	M	M	H	M	H	M

Weight of Evidence Overall Judgement Rating

An average rating was given for all studies, either: **Low** (L), **Medium-Low** (M-L), **Medium** (M), **Medium-High** (M-H), or **High** (H)

Criteria used for this review	Berkowitz et al. (2015)	Hiefje et al. (2017)	Kosko and Ferdig (2016)	Outhwaite et al. (2017) Study 1	Outhwaite et al. (2017) Study 4	Outhwaite et al. (in press)	Park et al. (2016)	Pitchford (2015)	Schacter and Jo (2016)	Schacter et al. (2016)	Schacter and Jo (2017)	van der Ven et al. (2017)
WoE Rating A	M	H	L	M	H	H	M	M	M	M	M	M
WoE Rating B	M	M	M	M	L	M	M	H	M	M	M	H
WoE Rating C	L	L	L	M	H	H	M	M	H	M	H	M
Overall Rating	M-L	M	L	M	M-H	H	M	M-H	M-H	M	M-H	M-H

8.4 Systematic Literature Review: Comparison of Included Study Features

A systematic map of the key features and findings extracted from each study included within the systematic review is presented in this appendix. Effect sizes are also reported in the table for any statistically significant findings. They are assigned a categorical label small, medium or large are assigned according to the rules of thumb stated by Cohen (1988):

Cohen's *d* and Hedge's *g*: small (0.2 – 0.5), medium (0.5 – 0.8), large (> 0.8)

Author (date) and Country	Participants	Design	Intervention	Mathematics Outcome Measure(s)	Key Findings	Effect Size (Significant Results Only)	Weight of Evidence Ratings
Berkowitz et al. (2015) USA	<i>n</i> = 587 6-7 years	Cluster RCT Comparison group using reading based app	Home-based intervention, with a parent acting as facilitator and using the apps with their children <i>Bedtime Learning Together</i> containing a maths passage followed by corresponding questions. Covers topics including geometry, arithmetic, fractions, counting and probability Several sessions per week over the course of a school year (Total usage monitored for each parent-child dyad)	Standardised measure: Woodcock-Johnson-III Tests of Achievement	The more times that parents and children used the app the greater children's maths achievement at the end of the intervention. However this pattern did not hold for the reading comparison group.	No effect size reported	A: Medium B: Medium C: Low Overall: Medium-Low

Author (date) and Country	Participants	Design	Intervention	Mathematics Outcome Measure(s)	Key Findings	Effect Size (Significant Results Only)	Weight of Evidence Ratings
Hieftje et al. (2017) Mexico	$n = 134$ 6-7 years	RCT Comparison group used non-maths games on an iPad tablet	Children used the intervention during after-school programmes. The type of facilitator and nature of their involvement was not reported <i>Knowledge Battle</i> : mini-games focused on the mathematics curriculum in Mexico 2-3 sessions per week over 4 weeks, each lasting 60 minutes (8-10 hours total gameplay)	Standardised Measure: KeyMath-3 Diagnostic Assessment	Significant intervention effect reported for the Numeration scale of the KeyMath-3 Diagnostic assessment, but not overall or on subscales assessing other aspects of maths, including measurement, addition/subtraction and problem solving However, a significant intervention effect was found for a lower achieving subgroup on overall maths scores and on the numeration subscale.	Cohen's d effect size not reported	A: High B: Medium C: Medium Overall: Medium-High
Kosko and Ferdig (2016) USA	$n = 73$ 3-5 years	RCT Treatment as usual control	Children used the intervention at home, with variable levels of support from their parents <i>Zorbit</i> : Contains six levels or worlds and children progress though to earn enough stars to power a rocket ship. Tasks focus upon number recognition, sorting/matching, counting, quantity comparison, understanding of ordinal numbers, spatial reasoning and geometry At least a weekly basis for three weeks	Researcher developed measure: 19 item test measuring quantitative reasoning, patterns, algebraic reasoning, geometry, spatial reasoning and arithmetic	Authors report a statistically significant effect of the intervention on post test mathematics ($p = 0.58$), but the effect was not statistically significant at a $p < .05$ level. There was however a statistically significant intervention effect on spatial reasoning aspects of the assessment.	Cohen's d effect size not reported	A: Low B: Medium C: Low Overall: Low

Author (date) and Country	Participants	Design	Intervention	Mathematics Outcome Measure(s)	Key Findings	Effect Size (Significant Results Only)	Weight of Evidence Ratings
Outhwaite, Gulliford and Pitchford (2017) Study 1 UK	$n = 83$ Intervention group: 4-5 years Control group: 5-7 years	Quasi-experimental Treatment as usual control	Children used the intervention in the classroom with the support of a member of teaching staff <i>Maths Age 3-5</i> and <i>Maths Age 4-6</i> apps developed by <i>onebillion</i> : tasks focus upon number, shape, space and measure concepts in the UK National Curriculum 5 sessions per week for 6 weeks, each lasting 30 minutes	Researcher developed measures: Curriculum Knowledge and Maths Concepts assessments (tablet administered)	Experimental group showed significant increases in curriculum knowledge and maths concepts following the intervention. At pre-test, an older control group achieved significant higher than the younger experimental group on maths concepts and curriculum knowledge. However at post-test and 5 month follow-up, the younger experimental group achieved a higher mean curriculum knowledge score than the older control children (although not statistically significant). No significant relationship between SES or EAL status and mathematics gains.	Within group effect sizes: Cohen's $d = 1.0$ (curriculum knowledge) = large Cohen's $d = 0.3$ (maths concepts) = small	A: Medium B: Medium C: Medium Overall: Medium

Author (date) and Country	Participants	Design	Intervention	Mathematics Outcome Measure(s)	Key Findings	Effect Size (Significant Results Only)	Weight of Evidence Ratings
Outhwaite, Gulliford and Pitchford (2017) Study 4 UK	$n = 27$ 4-5 years Low-attaining	Quasi-experimental Control: time-equivalent exposure to mathematics	Children used the intervention in the classroom with the support of a member of teaching staff 2 apps - <i>Maths Age 3-5</i> and <i>Maths Age 4-6</i> apps developed by <i>onebillion</i> : tasks focus upon number, shape, space and measure concepts in the UK National Curriculum Access on 50% of teaching days for 16 weeks, each lasting 30 minutes. (Equivalent to 8 weeks exposure, 5 sessions per week)	Standardised measure: Mathematical Reasoning sub-test from the WIAT-II Researcher developed measure: Curriculum knowledge test	Significantly greater gains on curriculum knowledge and maths concept knowledge made by low attaining children receiving intervention compared to higher attaining control children. No significant relationship between SES, memory and/or mathematics gains.	Within group effect sizes: Cohen's $d = 3.3$ (curriculum knowledge) = large Cohen's $d = 2.5$ (maths concepts) = large	A: High B: Low C: High Overall: High

Author (date) and Country	Participants	Design	Intervention	Mathematics Outcome Measure(s)	Key Findings	Effect Size (Significant Results Only)	Weight of Evidence Ratings
Outhwaite et al. (in press) UK	$n = 389$ 4-5 years	RCT Treatment as usual control	Children used the intervention in the classroom with the support of a member of teaching staff <i>Maths Age 3-5</i> and <i>Maths Age 4-6</i> apps developed by <i>onebillion</i> : tasks focus upon number, shape, space and measure concepts in the UK National Curriculum 5 sessions per week for 12 weeks, each lasting 30 minutes	Standardised measure: Progress Test in Maths, Level 5	Significant intervention effect on mathematics attainment for an experimental group using the maths app as well as their typical practice and for a group using the app instead of one regular mathematics activity. For a sub-set of low attaining children, learning gains were much greater for children who accessed the maths apps in addition to their typical curriculum compared to the control group	Between group effect sizes (progress over time): Cohen's $d = 0.31$ (as well as group) = small Cohen's $d = 0.21$ (instead of group) = small Within group effect size: Cohen's $d = 4.03$ (low attaining children – as well as group) = large Cohen's $d = 1.25$ (treatment as usual group)	A: High B: Medium C: High Overall: High

Author (date) and Country	Participants	Design	Intervention	Mathematics Outcome Measure(s)	Key Findings	Effect Size (Significant Results Only)	Weight of Evidence Ratings
Park, Bermudez, Roberts and Brannon (2016) Spain	$n = 103$ 3-5 years	RCT Comparison group used tablet-based picture memory activity	Children used the app in small groups, the nature/role of the facilitator was not specified Approximate arithmetic training: tasks focus upon addition and subtraction of arrays of objects 10 sessions over 2-3 weeks, each lasting 10-12 minutes	Standardised Measure: Test of Early Mathematical Achievement-Third Edition (TEMA-3)	Significant intervention effect reported for whole sample. Sub-sample analysis showed a significant intervention effect for lower income children and younger children < 4.9 years, but not older children.	Between group effect size: Cohen's $d =$ 0.414 (whole sample) = small	A: Medium B: Medium C: Medium Overall: Medium

Author (date) and Country	Participants	Design	Intervention	Mathematics Outcome Measure(s)	Key Findings	Effect Size (Significant Results Only)	Weight of Evidence Ratings
Pitchford (2015) Malawi	$n = 283$ Two groups according to educational year group Standard 1: 6 – 9 years Standard 2: 6 – 9 years	RCT Treatment as usual control Comparison group had access to non-mathematics tablet apps (equivalent time exposure)	Children used the intervention in a designated area with the support of a member of teaching staff 4 apps - <i>Masumu</i> (Chichewa for Maths) 1, <i>Masumu 2</i> , <i>Count to 10</i> and <i>Count to 20</i> developed by <i>onebillion</i> following the National Primary Curriculum in Malawi Standard 1: Alternate school days for 8 weeks, each lasting 30 minutes Standard 2: Alternate school days for 8 weeks, each lasting 60 minutes	Researcher developed measures: Mathematics Curriculum Knowledge and Mathematics Concepts (tablet administered) Maths Curriculum Knowledge Generalisation – post-test only	No significant intervention effect for children in educational Standard 1 for curriculum knowledge and mathematics concepts. For Standard 2 children, there was a significant intervention effect for mathematics concepts compared to normal practice but not non-maths tablet comparison groups. Significant intervention effect for curriculum knowledge compared to normal practice and non-maths tablet comparison groups. At post-test on the curriculum knowledge generalisation assessment the experimental group performed significantly better than the non-maths tablet comparison group but not the normal practice control group.	Between group effect sizes (post-test): Cohen's $d = 0.626$ (maths concepts) = medium Cohen's $d = 1.119$ (curriculum knowledge) = large Cohen's $d = 0.354$ (curriculum knowledge generalisation) = small	A: Medium B: High C: Medium Overall: Medium-High

Author (date) and Country	Participants	Design	Intervention	Mathematics Outcome Measure(s)	Key Findings	Effect Size (Significant Results Only)	Weight of Evidence Ratings
Schacter and Jo (2016) USA	$n = 162$ 4-5 years	Quasi-experimental Treatment as usual control	Children used the intervention in the classroom with the support of a member of teaching staff for the first two weeks, and then independently for the remaining weeks <i>Math Shelf</i> : tasks focus upon subitizing, counting, comparison, sequencing, number recognition and place value 2 sessions per week for 15 weeks, each lasting 10 minutes	Researcher developed measure: Number sense measure (tablet administered) testing quantity discrimination, numeral identification, numeral sequencing, cardinal principle, comparing quantities and matching numerals to quantities	Significant intervention effect for number sense The effect of the intervention was greater for children with lower pre-test scores.	Between-group effect size (progress over time): Cohen's $d = 1.09$ = large	A: Medium B: Medium C: High Overall: Medium-High
Schacter et al. (2016) USA	$n = 86$ 4-5 years	RCT Comparison group used other tablet mathematics apps	Children used the intervention in a separate classroom, supervised by a graduate student from the research team <i>Math Shelf</i> : tasks focus upon subitizing, counting, comparison, sequencing, number recognition and place value 3 sessions per week for 6 weeks, each lasting 10 minutes	Researcher developed measure: Mathematics assessment (tablet administered) testing quantity discrimination, numeral identification, numeral sequencing, cardinal principle, comparing quantities and matching numerals to quantities	Significant intervention effect for number sense. The effect of the intervention was greater for children with higher pre-test scores and for female students.	Between group effect size (progress over time): Cohen's $d = 0.57$ = medium	A: Medium B: Medium C: Medium Overall: Medium

Author (date) and Country	Participants	Design	Intervention	Mathematics Outcome Measure(s)	Key Findings	Effect Size (Significant Results Only)	Weight of Evidence Ratings
Schacter and Jo (2017) USA	$n = 378$ 4-5 years	RCT Comparison group used other tablet mathematics apps	Children used the intervention in the classroom with the support of a member of teaching staff for the first two weeks, and then independently for the remaining weeks <i>Math Shelf</i> : tasks focus upon subitizing, counting, comparison, sequencing, number recognition and place value 2 sessions per week for 22 weeks, each lasting 10 minutes	Researcher developed measure: Mathematics assessment (tablet administered) testing numeral identification, cardinal principle, numeral sequencing, matching numerals to quantities, quantity discrimination, place value and addition	There was a significant intervention effect on numeracy knowledge. The effect of the intervention was greatest for students with lower pre-test numeracy scores.	Between group effect size (progress over time): Cohen's $d = 0.94$ = large	A: Medium B: Medium C: High Overall: Medium-High
van der Ven et al. (2017) Netherlands	$n = 103$ 6-7 years	RCT Treatment as usual control	Children used the intervention, supervised by a member of teaching staff. A racing game to develop arithmetic efficiency with addition and subtraction problems Four sessions per week for 5 weeks, each lasting 15 minutes.	Researcher developed measure: arithmetic fluency test	There was a significant intervention effect on children's fluency in solving dot-subtraction problems, but not problems involving Arabic symbols or dot-addition. The effect was not significant at longitudinal follow-up.	Cohen's d not reported	A: Medium B: High C: Medium Overall: Medium-High

8.5 Expression of Interest Letter



The University of
Nottingham

UNITED KINGDOM • CHINA • MALAYSIA

Faculty of Science
School of Psychology

The University of Nottingham
University Park

Nottingham

NG7 2RD

t: +44 (0)115 951 5361

e: psychology@nottingham.ac.uk

www.nottingham.ac.uk/psychology

(Insert date)

Dear (Insert name of head teacher)

I am currently working as a trainee Educational Psychologist in [town] supervised by [name]. As part of my doctoral training at the University of Nottingham, I am hoping to conduct a research project in [county] schools to evaluate the impact of tablet technology on mathematics attainment in young children. I am contacting you to ask whether your school might be interested in participating in this research.

The charity *onebillion* have designed a series of mathematics apps for children aged 3-6 years, based on the UK curriculum. The apps contain a series of activities, modelled by a virtual teacher, which children progress through at their own pace and receive instant feedback for their responses. Previous research conducted in Malawi and in the UK with Year 1/FS2 has shown that the apps can lead to significant improvements in children's mathematical achievement, and are particularly effective for low-attaining children. The apps therefore show promise in helping to 'close the gap' in mathematical attainment at an early age.

However, to date, research has not looked at whether the intervention is effective for slightly younger children, aged 3-4 years. This would be the focus of my study, which I hope would run over the summer term of this year with children in nursery/Foundation Stage 1 classes. The main requirement for participating schools is that children in these classes can have regular access to Apple iPad tablet devices (minimum of 10).

The benefits of participating in this study would be as follows:

- Opportunity for independent evaluation of the use of technology within school for raising standards in mathematics

- Access to standardised assessment data about the mathematics attainment of participating pupils to inform future teaching
- Free access to the software, which previous research has shown to be effective for children in FS2/Year 1, and particularly for low-attaining children.
- Opportunity to share with parents/carers how the school is using technology to enhance learning
- Evidence to Ofsted that schools are taking a reflective approach to their teaching and supporting wider educational research.

At the moment I am not expecting any formal agreement to take part in the study, nor will an expression of interest at this stage guarantee that you will be selected to participate. However, I would be grateful if you could contact me by the end of the autumn term if you would be interested in finding out more about the proposed research. Please also do not hesitate to contact me, [name] (Placement Supervisor) or Anthea Gulliford (Research Supervisor), if you have any further queries using the details below.

I look forward to hearing from you,

Best Wishes

Jodie Walton

Trainee Educational Psychologist

Contact Details:

Jodie Walton

[name] (Placement Supervisor)

Anthea Gulliford (Research Supervisor)

8.6 Head Teacher Information Sheet and Consent Form



The University of
Nottingham

UNITED KINGDOM • CHINA • MALAYSIA

Information Sheet for Schools

An evaluation of the impact of a tablet-based intervention on the mathematics attainment, receptive language skills and ‘approach to learning’ of 3-4 year old children

Ethics Approval Number: S938

Researcher: Jodie Walton

University Research Supervisor: Anthea Gulliford

Placement Supervisor: [name]

Contact Details:

Jodie Walton (Trainee Educational Psychologist and Researcher) at

Email:

Tel:

Anthea Gulliford (Research Supervisor) at

Email:

Tel:

[name] (Placement Supervisor) at

Tel:

This is an invitation for your school to take part in a research study exploring the use of tablet technology to improve young children’s mathematics skills.

Before you decide if you wish for your school 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.

The purpose of the study is to evaluate a mathematics app, *Maths Age 3-5*, developed by a not-for-profit organisation, onebillion. Previous research has shown that this the app can improve the mathematics attainment of 4-6 year old children. This study aims to find out whether the intervention is beneficial for slightly younger children, aged 3-4 years. The research also aims to find out whether the intervention has any impact upon children’s understanding of language and the way that they approach learning tasks (e.g. their persistence in problem solving).

The app provides tuition across several key topics in mathematics, covering content outlined within the UK National Curriculum for this age range. A virtual

teacher delivers the 'lessons' in an accessible and attractive format. Children receive feedback for their responses and can progress through activities at their own pace. Children complete quizzes at the end of each topic, which allows their progress to be monitored by staff.

If you agree for the school to be involved, then the research will involve the participation of children and teaching staff in Foundation Stage 1. Parental and pupil consent must be freely obtained for each child involved in the research, and the researcher will pass these letters to school for distribution. Staff must also freely give consent to participate in the research.

The project will begin with an individual assessment of all children's mathematics and language skills, using short activities and games. Assessments will be conducted by the researcher over a series of three sessions, each lasting no longer than 15 minutes, and will require a quiet room close to the child's classroom. The nursery class teacher will also be asked to complete a brief questionnaire about each child's 'approach to learning', which is expected to take 2-3 minutes per child. Questionnaires can be completed over a period of a week.

Children will then be randomly allocated to one of the following groups:

- An intervention group, who will use the app over the summer term for one 15 minutes session each day whilst the other children engage in usual play-based activities. Children will not miss any adult-directed teaching sessions in mathematics or any other area of learning.
- A control group who will continue to receive their normal mathematics teaching. These children will be able to access the intervention in school at the end of the project when they enter Foundation Stage 2 if they remain at the same setting.

Each intervention group will require supervision by a member of staff, who will receive training prior to the start of the project and ongoing support throughout the intervention. The training will take no longer than an hour, to explain and discuss how the app can best be implemented in the classroom. The intervention will last for 9 weeks and then children in all groups will be re-assessed.

After the intervention, an interview(s) will then be conducted with the member(s) of staff who facilitated the intervention, subject to their consent. This is expected to last no longer than 30 minutes. The purpose of the interview is to find out staff views on how the intervention has gone, and how well it has worked.

Children will then resume normal mathematics teaching for four months and will be assessed again when they are in Foundation Stage 2 (December) in order to

evaluate whether the intervention is effective over a longer time period. After this point, all children will have access to the intervention during school time, including children previously allocated to the control group.

Assessment sessions will be undertaken by the researcher, who is Trainee Educational Psychologist within [name] Local Authority, and is also a qualified primary school teacher. The school will be provided with a full Disclosure and Barring Services (DBS) clearance certificate for the researcher.

Participation in this study is totally voluntary and you are under no obligation to take part. You and your school are free to withdraw at any point before or during the study. This will not affect your right to access other services provided by the Educational Psychology Service.

All data collected will be kept confidential and used for research purposes only. It will be stored in compliance with the Data Protection Act. School staff will be able to access non-anonymised data about the mathematics attainment, language skills and 'approach to learning' of pupils. Parents will also be able to request access to the data for their child. Data will be anonymised in any research reports or outputs.

The findings of the research will be shared with your school and other audiences through various summaries, to help schools to develop their use of educational technology and mathematics instruction. Parents/carers will also receive a summary of the key findings of the research.

Benefits to the School

- Opportunity for independent evaluation of the use of technology within school for raising standards in mathematics.
- Access to standardised assessment data about the mathematics attainment of participating pupils to inform future teaching.
- Free access to the software, which previous research has shown to be effective for children in FS2/Year 1, and particularly for low-attaining children.
- Opportunity to share with parents/carers how the school is using technology to enhance learning.
- Evidence to Ofsted that schools are taking a reflective approach to their teaching and supporting wider educational research.

Roles and Responsibilities

The Researcher

- To host an information evening for parents/carers about the project and to provide letters for schools to send out to parents/carers to inform them about the meeting.

- To explain the intervention to staff and provide training in its implementation
- To administer measures of attainment and provide questionnaires to school staff at each time point.
- To observe four sessions of the intervention in the summer term.
- To contact the facilitator weekly to monitor the intervention and address any concerns
- To ensure that the control group receive the intervention in the Spring Term of Foundation Stage 2, and that staff are provided with training.
- To ensure all data is fully anonymised and held confidentially, in line with the Data Protection Act
- To feedback the findings of the research to parents/carers, nurseries, the Local Authority and to debrief all participants and stakeholders.

The School

- To nominate a member of staff as a point of contact for the researcher and parents/carers.
- Allow the researcher to host an information evening at the school for parents in order to obtain their informed consent, and to send out information about the meeting to parents/carers.
- Send out information letters and consent forms to parents, prepared by the researcher, and collect forms returned to the school.
- Allow the researcher to individually assess each pupil involved in the research (pre, post and 4 months after the end of the intervention), providing a quiet room close to the children's classroom. It will be useful to allocate a member of staff to introduce Jodie to each of the children before testing.
- To allow the nursery class teacher time to complete questionnaires about each child's approach to learning before and after the intervention.
- Ensure the availability of iPad tablet devices and headphones for one 15 minute session per day for all children involved in the research over a 9 week period (either in the summer of FS1 for children in the intervention group, or spring of FS2 for children in the control group)
- To release nursery staff for 30 minutes at the end of the summer term so that they can be interviewed about the project.

If you have any questions or concerns please don't hesitate to ask now. We can also be contacted after your participation at the above address.

If you have any complaints about the study, please contact:
 Stephen Jackson (Chair of Ethics Committee)
stephen.jackson@nottingham.ac.uk



Head Teacher Consent Form

An evaluation of the impact of a tablet-based intervention on the mathematics attainment, receptive language skills and 'approach to learning' of 3-4 year old children

Ethics Approval Number: S938

Researcher: Jodie Walton

University Research Supervisor: Anthea Gulliford

The head teacher should answer these questions independently:

- Have you read and understood the Information Sheet?
YES/NO
- Have you had the opportunity to ask questions about the study?
YES/NO
- Have all your questions been answered satisfactorily?
YES/NO
- Do you understand that you are free to withdraw from the study? (at any time and without giving a reason)
YES/NO
- I give permission for school data from this study to be shared with other researchers provided that anonymity is completely protected.
YES/NO
- Do you agree to take part in the study?
YES/NO

"This project has been fully explained to me and I agree that our school will take part. I understand that I have the right to withdraw at any time without giving a reason and that this will not affect access to the Educational Psychology Service. I understand that parental and pupil consent must be obtained for each child taking part in the research."

Signature of head teacher:

Date:

Name (in block capitals)

I have explained the study to the above head teacher and he/she has agreed that their school will take part.

Signature of researcher:

Date:

8.7 Parent Information Sheet and Consent Form



The University of
Nottingham

UNITED KINGDOM • CHINA • MALAYSIA

Information Sheet for Parents/Carers

An evaluation of the impact of a tablet-based intervention on the mathematics attainment, receptive language skills and 'approach to learning' of 3-4 year old children

Ethics Approval Number: S938

Researcher: Jodie Walton

University Research Supervisor: Anthea Gulliford

Placement Supervisor: [name]

Contact Details:

Jodie Walton (Trainee Educational Psychologist and Researcher) at

Email:

Tel:

Anthea Gulliford (Research Supervisor) at

Email:

Tel:

[name] (Placement Supervisor) at

Tel:

Dear Parent

I am currently training as an Educational Psychologist in [county], after previous experience as a primary school teacher. As part of my doctoral training at the University of Nottingham, I am conducting a research project at your child's nursery, exploring the use of tablet technology to improve young children's mathematics skills. I am contacting you to ask your permission for your child to take part in this research.

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.

The purpose of the study is to evaluate a mathematics app, *Maths Age 3-5*, developed by a not-for-profit organisation, onebillion. Previous research has shown that this app can improve the mathematics skills of 4-6 year old children. This study aims to find out whether the intervention is beneficial for

slightly younger children, aged 3-4 years. The research also aims to find out whether the intervention has any impact upon children's understanding of language and the way that they approach learning tasks.

If you agree for your child to participate, their mathematics and language skills will be individually tested by the researcher at the start of the project, using short activities and games. Assessments will take place over three short sessions, each lasting no longer than 15 minutes, in a quiet room close to the child's classroom. The class teacher will also complete a short questionnaire about the way that your child approaches learning tasks in the nursery.

Children will then be randomly placed in one of the following groups:

Group A: Children in this group will use the app for 15 minutes per day for 9 weeks during the summer term. This will happen whilst the other children are involved in play activities so that they do not miss any whole-class teaching sessions in mathematics or any other area of learning. They will work independently in small groups under the supervision of school staff. Activities within the app include, sorting, matching, counting and understanding patterns, which are aligned to the activities they are being taught in class.

Group B: Children in this group will continue to receive their normal mathematics teaching in the summer term. These children will be able to access the app in school at the end of the project (in the Spring Term of Foundation Stage 2), if they remain at the same school setting.

The intervention will last for 9 weeks and then children in all groups will be re-assessed.

In the following December, when the children are in Foundation Stage 2, their mathematics skills and knowledge will be assessed again to see whether the app has a lasting benefit. If your child will be attending a different school in the autumn term, then you will be contacted by letter to give details of your child's new school setting (if you are happy for them to be re-assessed). If the school is within [county], then the researcher will be able to conduct the assessment at the child's new school in December if the head teacher agrees.

Participation in this study is totally voluntary and you are under no obligation to allow your child to take part. You and your child are free to withdraw at any point before or during the study. This will not affect your right to access other services provided by the Educational Psychology Service. All data collected will be kept confidential and used for research purposes only. Your child will not be identified by name in any research summaries. Data will be stored in compliance with the Data Protection Act. If you choose to provide your postcode it will be used as an estimate of socio-economic status to determine whether the app is beneficial to all children.

If you have any questions or concerns please don't hesitate to ask. We can also be contacted after your participation using the details above.

If you have any complaints about the study, please contact:
Stephen Jackson (Chair of Ethics Committee)
stephen.jackson@nottingham.ac.uk



An evaluation of the impact of a tablet-based intervention on the mathematics attainment, receptive language skills and ‘approach to learning’ of 3-4 year old children

Ethics Approval Number: S938
Researcher: Jodie Walton (E-mail:)
Supervisor: Anthea Gulliford (Email:)

The parent/carer should answer these questions independently:

- Have you read and understood the Information Sheet?
YES/NO
- Have you had the opportunity to ask questions about the study?
YES/NO
- Have all your questions been answered satisfactorily?
YES/NO
- Do you understand that you are free to withdraw your child from the study? (at any time and without giving a reason)
YES/NO
- I give permission for my child’s data to be shared with other researchers provided that their anonymity is completely protected.
YES/NO
- Do you agree for your child to take part in the study?
YES/NO

“This study has been explained to me to my satisfaction, and I agree to take part. I understand that I am free to withdraw my child at any time and that this will not affect access to the Educational Psychology Service.”

Name of Child:

Postcode:

Child's Date of Birth:

Child's First Language²¹:

Signature of Parent/Carer:

Date:

Name (in block capitals):

I have explained the study to the above parent/carer and he/she has agreed for their child to take part.

Signature of researcher:

Date:

²¹ *First language = The main language your child has encountered as a baby or small child. They do not have to be fully fluent in this language.*

8.8 Child 'Willingness to Participate' Form



The University of
Nottingham

UNITED KINGDOM • CHINA • MALAYSIA

Child 'Willingness to Participate' Form An evaluation of the impact of a tablet-based intervention on the mathematics attainment, receptive language skills and 'approach to learning' of 3-4 year old children

Ethics Approval Number: S938

Researcher: Jodie Walton

University Research Supervisor: Anthea Gulliford

Placement Supervisor: [name]

Child's ID:
(Read aloud to child)

Date:

Hello (Child's name),

My name is (own name), and I am a trainee psychologist. My job is to visit different schools and help children with their learning. I am working with the children in your class at the moment to help them to learn maths and you can take part as well if you would like to. Today, I would like to do some different maths activities and listening games with you. If you do not want to answer any questions or wish to stop taking part then that is fine and I will take you back to your classroom.

Do you have any questions? (Answer child's questions)

Would you like to do these activities? (tick)

☐☐☐☐☐☐

Tell me if you change your mind because we can stop at any time.

I have explained the study to the pupil named above and he/she has agreed to take part.

Signature of researcher:

Date:

8.9 Curriculum Knowledge Assessment Example of Questions and Answer Booklet (from Outhwaite et al., 2017)

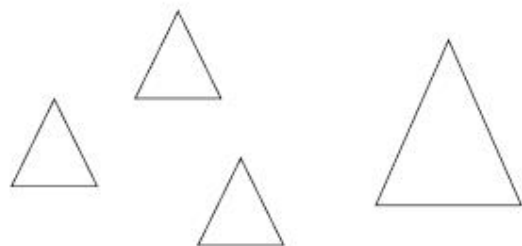
*Onebillion maths test*²²

Maths test Version (First six questions)

	Question	Answer	Notes
1.	Which is the odd one out?	Big triangle (marks/points)	
2.	Which is the matching pair?	2 owls (marks/points)	
3.	Show me 2 stars	Colours/marks/ points to 2 stars	
4.	How many frogs are there?	4 (verbal or written)	
5.	What comes next in the pattern?	Circle (verbal or written)	
6.	Where are the triangle and the square?	Marks/points to the triangle and square	1 point for both. No points if only one shape is identified.

²² Presented within this thesis with permission from the authors

1.



2.



3.



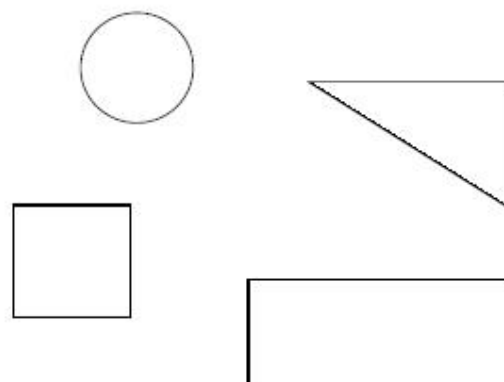
4.



5.



6.



8.10 Facilitator Training

What is the *onebillion* mathematics tablet intervention?

The *onebillion* tablet intervention involves two separate apps, *Maths Age 3-5* and *Maths Age 4-6*. Both of these apps follow topics in the National Curriculum for Foundation Stage. A 'virtual' teacher delivers the 'lessons', providing clear modelling and instructions for each topic area, followed by opportunities for independent practice. Content in the apps has been unlocked on a selection of school iPad tablet devices– please talk to nursery staff about which tablets to use.

A full list of the topics on the apps and a breakdown of the activities is available on the *onebillion* website.

<https://onebillion.org/apps/maths3to5>

<https://onebillion.org/apps/maths4to6>

Within each topic area there are ten different activities for the children to complete to practice the skill, followed by a short quiz. Each child has their own profile on the app so that they can progress through the activities at their own pace and learning is personalised for them. Once children have completed the quiz and answered each question correctly they receive a virtual certificate.

How did the intervention run last time?

- Last time the intervention ran for 9 weeks, for 15 minutes a day. Children used the intervention at the same time each day so that it became part of the class routine.
- Children accessed the apps in small groups, supervised by a teacher. They wore headphones to listen to the virtual teacher's instructions.
- Last time all children began from the beginning of the first app and then progressed through activities from the same point that they reached the last day. The app remembers the activity each child last completed, prompting children to start from the activity that flashes. You may wish for some children to start at a later point in the intervention now, depending on teacher assessment. However, bear in mind that the app activities may help to consolidate prior learning that has already been taught in class before children progress on.

What is the role of teaching staff?

- The app does not replace the need for some teaching support, although this will depend on the needs of each individual child.
- Teaching staff should support children to work independently where appropriate, but provide additional teaching support so that children make progress.

Support varies depending upon the child's need but might include:

- Providing praise, encouragement and prompts to maintain attention
- Prompting the child to press the ear in the top right corner to listen again to the instructions
- Removing headphones and providing additional teaching support where needed

Important points to note

- **Tracking progress:** I have sent you a tracking grid so that you can monitor children's progress through the app. There are two trackers, one for each app.
- **Quizzes:** If children don't pass a quiz, then consider asking them to go back and practice some of the earlier activities before trying again. However, after a few turns, provide them with support to complete the quiz if necessary so that they don't become discouraged. They can then move onto the next topic area, which may be a different area of maths that they can succeed at.
- **Monitoring progress:** Take care to monitor children's progress carefully, provide extra support and try a different teaching approach instead if necessary so that children don't become frustrated or discouraged. If any child experiences persistent distress, despite additional teaching support, they should be removed from the tablet group.
- **Freezing:** Unfortunately there is a 'bug' in one of the activities – the 'odd one out' activity towards the beginning of the Maths Age 3-5 app, where children are presented with a selection of objects and have to draw a line through the odd one out. The app sometimes freezes on this activity. The best way to overcome freezing is to remind children to simply tap on the odd one out object rather than cross it out. However, you could skip this activity if it is easier. I have been informed that the developers at *onebillion* are working to remove this bug at the moment.

8.11 Intervention Fidelity Check

Date:

School:

Group:

Facilitator:

Fidelity Checklist

Feature	Evaluation Criteria	Y/N (Time)	Additional Comments/Feedback given to Facilitator
Intervention lasts for 15 minutes	Teacher stops children 15 minutes after all of the children have logged into the app (15 mins+/- 30 seconds)		
			Actual Time taken:
Children access their own profile on the app	All children use their own profile on the app or are redirected to their own profile within the first 3 minutes of the intervention		
Children progress through the app from the first activity in a sequence	All children begin using the app from the next activity in the sequential progression suggested by the app OR children repeat activities that they have already completed for additional practice if they have not passed a test		
Children wear headphones	All children wear headphones during the use of the app, or wear their headphones again within 30 seconds of removing them		

Diffusion of Treatments

Control children have not accessed intervention (Teacher report)	Yes/No	Comments:
--	--------	-----------

Control Group

All children continue to receive their typical mathematics instruction (Teacher report)	Yes/No	Comments:
---	--------	-----------

8.12 Facilitator Information Sheet and Consent Form



The University of
Nottingham

UNITED KINGDOM • CHINA • MALAYSIA

Information Sheet for Nursery Facilitators

An evaluation of the impact of a tablet-based intervention on the mathematics attainment, receptive language skills and 'approach to learning' of 3-4 year old children

Ethics Approval Number: S938

Researcher: Jodie Walton

University Research Supervisor: Anthea Gulliford

Placement Supervisor: [name]

Contact Details:

Jodie Walton (Trainee Educational Psychologist and Researcher) at

Email:

Tel:

Anthea Gulliford (Research Supervisor) at

Email:

Tel

[name] (Placement Supervisor) at

Tel:

This is an invitation for you to take part in a research study exploring the use of tablet technology to improve young children's mathematics skills.

As you will be facilitating the intervention in your nursery, I would like to invite you to take part in this research. Before you decide if you wish 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.

The purpose of the study is to evaluate a mathematics app developed by the charity, onebillion. Previous research has shown that this the app can improve the mathematics attainment of 4-6 year old children. This study aims to extend these findings by determining whether the intervention is effective for slightly younger children, aged 3-4 years. The research also aims to determine whether the intervention has any impact upon children's understanding of language and the way that they approach learning tasks (e.g. their persistence and independence).

If you agree to participate, then I will ask you to complete short questionnaires for each child participating in the project, which will evaluate their approach to

learning tasks in the nursery. The questionnaires will take 2-3 minutes to complete for each child and you will be able to complete the questionnaires over a period of a week. I will ask you to complete the questionnaires at the end of the spring term, and then again at the end of the summer term once the intervention has been completed. This will allow me to determine whether the intervention affected the way that children approach learning tasks.

After the intervention is complete, I would like to interview you to discuss your views about the outcomes of the intervention and the factors that may have affected these outcomes. It is hoped that this information might help us to understanding more about the benefits of the intervention and how the apps might be developed in the future. I will be interviewing teachers from each of the nurseries participating in the research and I will provide all schools with a summary of the findings. The interviews will be recorded but the information you give will be fully anonymised in the reporting of results.

Participation in this study is totally voluntary and you are under no obligation to take part. You are free to withdraw at any point before, during or after the study. This will not affect your right to access other services provided by the Educational Psychology Service. All data collected will be kept confidential and used for research purposes only. It will be stored in compliance with the Data Protection Act.

If you have any questions or concerns please don't hesitate to ask now. We can also be contacted after your participation at the above address.

If you have any complaints about the study, please contact:
Stephen Jackson (Chair of Ethics Committee)
stephen.jackson@nottingham.ac.uk



An evaluation of the impact of a tablet-based intervention on the mathematics attainment, receptive language skills and ‘approach to learning’ of 3-4 year old children

Ethics Approval Number: S938

Researcher: Jodie Walton

University Research Supervisor: Anthea Gulliford

Placement Supervisor: [name]

The teacher should answer these questions independently:

- Have you read and understood the Information Sheet?
YES/NO
-
- Have you had the opportunity to ask questions about the study?
YES/NO
-
- Have all your questions been answered satisfactorily?
YES/NO
- Do you understand that you are free to withdraw
from the study? (at any time and without giving a reason)
YES/NO
-
- I give permission for my data from this study to be shared with other
researchers provided that my anonymity is completely protected.
YES/NO
-
- Do you agree to take part in the study?
YES/NO

“This project has been fully explained to me and I agree to take part. I understand that I have the right to withdraw at any time without giving a reason and that this will not affect access to the Educational Psychology Service.”

Signature of participant:

Date:

Name (in block capitals)

I have explained the study to the above participant and he/she has agreed to take part.

Signature of researcher:

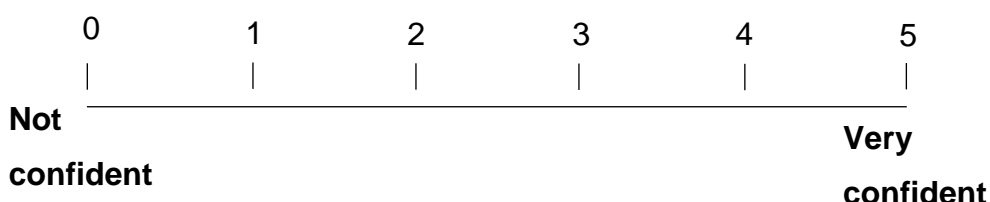
Date:

8.13 Semi-Structured Facilitator Interview Script

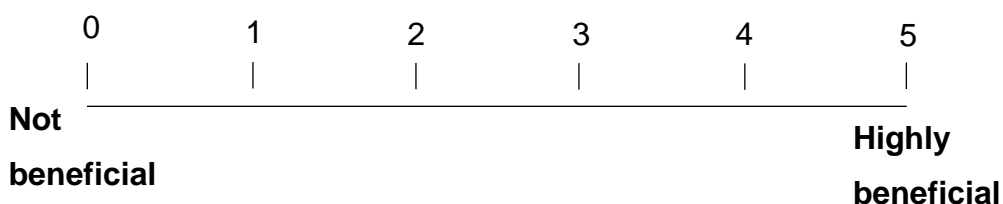
Thank you for agreeing to meet with me today following the end of the *onebillion* maths tablet intervention. I am interested in finding out your views about how the outcomes of the intervention and how well you feel it has worked for the children in your class. There are no right or wrong answers so please feel free to answer as honestly as you can. If you do not want to answer a questions then that is fine and we will move on to the next question. The interview will be recorded but all identifying information, including names of children and staff, will be anonymised in the reporting of the results. Please can you confirm that you are happy to take part and for this interview to be recorded?

I'll first ask you some background questions about your role in the school.

- 1.) What is your job title/role?
- 2.) How long have you been working in your current role?
- 3.) How long have you been working with children in the early years?
- 4.) Have received any qualifications that support your current role?
- 5.) Do you use a tablet at home yourself?
- 6.) Have you used tablets with children in school before this project?
(If yes) What have you used them for?
- 7.) Looking at this scale from 0-5, how would you rate your confidence in using tablet technology for early years' education?



- 8.) Looking at this scale from 0- 5, how beneficial do you feel that tablet technology can be within early years' education?



We'll now talk more specifically about the tablet intervention and how you feel that it went. We'll look at following key areas (Show overview of topics):
outcomes of the intervention, different children's responses to the intervention, your role as a facilitator, any additional classroom or school factors that might have affected outcomes, the features of the intervention itself and your future use of the app.

1.) Overall outcomes

What do you think were the outcomes of the tablet intervention for the children who were using the apps in your class? (Prompt for evidence: How do you know? In what way? What have you observed?)

Prompts if needed – Do you think there has been any impact on...

- Maths (if so, any specific aspects)?
- Language and listening skills?
- Approach to learning tasks (such as persistence, confidence, problem solving and independence)?
- Any other areas of learning or behaviour?

Were you happy with these outcomes? Why/why not?

2.) Individual differences in response to the intervention

Here is a list of children that participated in the intervention.

Did you notice any differences in the way that particular children responded to the intervention?

(If yes)

Which children do you feel benefited most from the intervention? Why?

Which children do you feel benefited less from the intervention? Why?

3.) Role of the facilitator

How independently were the children able to use the apps?

What support did you provide for the children whilst they were using the intervention?

(Prompts)

- Technical
- Learning and development

Did particular children need different levels or types of support? If so, why?

4.) Additional factors affecting outcomes

How easy did you find it to implement the intervention in your class? Why?

Were there any technical difficulties?

What aspects of the class environment, such as class groupings, staffing or

the activities of other children, may have affected the outcomes of the intervention?

What aspects of the school organisation and environment might have affected the outcomes of the intervention?

What other factors, if any, might have affected the outcomes of the intervention?

5.) Features of the intervention

What features of the intervention did you think were positive? Why?

What features of the intervention did you think were less beneficial? Why?

How do you think the *onebillion* tablet intervention compares to other types of adult-led maths teaching that the children receive?

How do you think the apps could be improved?

6.) Future use

Do you plan to continue using the *onebillion* apps in the future? Why? Why not?

If yes, will you use the apps differently to the way that they have been used in this project?

Is there anything further that you would like to say about your experience of implementing this intervention and its outcomes for children?

Give debriefing information

8.14 Debriefing Procedures

Assessments with Children

After each assessment the following script was used:

Thank you for completing these activities today. You worked really hard (*give child a sticker*) and I will share how you've worked today with your teacher.

First/Second Sessions: I will see you again tomorrow/soon to do some more activities if you would still like to take part.

Third session of time points 1 and 2: This will be the last time that I see you for a little while so thank you for working with me.

Final session at time point 3: This will be the last time that I will work with you, so thank you for taking part in the project.

Is there anything you would like to ask or tell me before you return to class?

Semi-structured interviews with teaching staff

The following debriefing statement on the next page was read to teaching staff following interviews.



An evaluation of the impact of a tablet-based intervention on the mathematics attainment, receptive language skills and 'approach to learning' of 3-4 year old children

Ethics Approval Number: S938

Researcher: Jodie Walton

University Research Supervisor: Anthea Gulliford

Placement Supervisor: [name]

Contact Details:

Jodie Walton (Trainee Educational Psychologist and Researcher) at

Email:

Tel:

Anthea Gulliford (Research Supervisor) at

Email:

Tel

[name] (Placement Supervisor) at

Tel:

Debriefing Information for Nursery Class Teachers

Thank you for taking the time to participate in this interview. The purpose of the interview was to give you an opportunity to share your views about the outcomes of the tablet-based maths intervention that you have facilitated with the nursery children. It is hoped that this information might help us to have a better understanding of the benefits of the intervention for young children and will inform the future development of the intervention.

Do you have any particular concerns about the intervention that you would like to discuss further or which you would like me to share with the head teacher?

I will now transcribe the data that has been recorded and ask you to check it for accuracy. I will then identify any themes that emerge across the viewpoints of different nursery staff and ask you whether you feel that your views have been interpreted correctly. The school will receive a final summary of the data. I will make sure that the data is anonymised in the reporting of the research. Please let me know if there are any particular comments that you would like me to remove from the recording.

I would like to remind you that you can still choose to withdraw and I can remove your data from the study at any stage. Please contact me, or my supervisors, using the details above if you have any concerns or you would like to discuss this interview further.

If you have any complaints about the study, please contact:
Stephen Jackson (Chair of Ethics Committee)
stephen.jackson@nottingham.ac.uk



8.15 Ethical Approval

SJ/wb
Ref:S938

Monday, 27 February 2017

Dear Jodie Walton & Anthea Gulliford,

Ethics Committee Review

Thank you for submitting an account of your proposed research 'An evaluation of the impact of a tablet-based intervention on the mathematics attainment, receptive language skills and 'approach to learning' of 3-4 year old children'.

That proposal has now been reviewed and we are pleased to tell you it has met with the Committee's approval.

However:

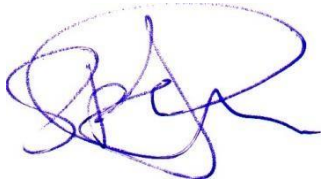
Please note the following comments from our reviewers;

The application could contain a more detailed description of the training exercises the children will be submitted to. In addition, given the young age of the participants and the fact that the experimenter is (at least initially) unfamiliar with the children, I was wondering if a second person (i.e., an employee of the nursery) should be present during the tests.

Final responsibility for ethical conduct of your research rests with you or your supervisor. The Codes of Practice setting out these responsibilities have been published by the British Psychological Society and the University Research Ethics Committee. If you have any concerns whatever during the conduct of your research then you should consult those Codes of Practice. The Committee should be informed immediately should any participant complaints or adverse events arise during the study.

Independently of the Ethics Committee procedures, supervisors also have responsibilities for the risk assessment of projects as detailed in the safety pages of the University web site. Ethics Committee approval does not alter, replace, or remove those responsibilities, nor does it certify that they have been met.

Yours sincerely

A handwritten signature in blue ink, appearing to be 'S. Jackson', enclosed within a faint, light blue oval border.

Professor Stephen Jackson Chair, Ethics Committee

8.16 Preliminary Checks for Analysis of Covariance

This appendix outlines the checks that were conducted on the data to determine whether each DV met the statistical assumptions required for ANCOVA analysis. Checks were conducted separately for experimental and control conditions.

Checks for Research Question 1 and 2

Assumption and Method of Testing	T2 ENC Ability Scores	T2 Mathematics CK Raw Scores	T3 Mathematics CK Raw Scores
1. Level of the Data	Interval level data (<i>Children cannot score an ability score of 0</i>)	Ratio level data	Ratio level data
2. Normality <ul style="list-style-type: none"> Visual analyses to explore Q-Q plots, identifying values deviating greatly from those expected in a normal distribution. Shapiro-Wilk tests to assess the null hypothesis that the data did not deviate significantly from normality. Normality can be assumed if $p > .05$ 	<ul style="list-style-type: none"> Q-Q plots indicated no clear deviation from normality in either condition. Shapiro Wilk tests were not significant for either condition. (Control $p = .994$; Intervention $p = .478$) <p>Normality was assumed.</p>	<ul style="list-style-type: none"> Q-Q plots indicated no clear deviation from normality in either condition. Shapiro Wilk tests were not significant for either condition. (Control $p = .230$; Intervention $p = .402$) <p>Normality was assumed.</p>	<ul style="list-style-type: none"> Q-Q plots indicated no clear deviation from normality in either condition Shapiro Wilk tests were not significant for either condition. (Control $p = .479$; Intervention $p = .578$) <p>Normality was assumed.</p>

Assumption and Method of Testing	T2 ENC Ability Scores	T2 Mathematics CK Raw Scores	T3 Mathematics CK Raw Scores
<p>3. Homogeneity of Variance</p> <ul style="list-style-type: none"> Levene's test to assess the null hypothesis that the variances from each sample do not differ significantly from each other. Homogeneity of variance can be assumed if $p > .05$ 	<p>The test was not statistically significant, $F(1, 41) = .620$, $p = .436$, and therefore homogeneity of variance was assumed.</p>	<p>The test was not statistically significant, $F(1,38) = .159$, $p = .692$, and therefore homogeneity of variance was assumed.</p>	<p>The test was not statistically significant, $F(1,34) = .050$, $p = .824$, and therefore homogeneity of variance was assumed.</p>
<p>4. Linear relationship between the DV and covariate</p> <ul style="list-style-type: none"> Scatterplots were constructed to visually analyse the relationship between both the DV and pre-test scores 	<p>Visual analysis of the scatterplot indicated an approximately linear relationship between the ENC Ability Scores at T1 and T2 and that this assumption was not violated.</p>	<p>Visual analysis of the scatterplot indicated an approximately linear relationship between the Mathematics CK Scores at T1 and T2 and that this assumption was not violated.</p>	<p>Visual analysis of the scatterplot indicated an approximately linear relationship between the Mathematics CK Scores at T1 and T3 and that this assumption was not violated.</p>

Assumption and Method of Testing	T2 ENC Ability Scores	T2 Mathematics CK Raw Scores	T3 Mathematics CK Raw Scores
<p>5. Homogeneity of regression across all experimental groups</p> <ul style="list-style-type: none"> Statistical testing to check that there was not a significant interaction between the treatment condition and covariate. Homogeneity of regression can be assumed if $p > .05$ 	Statistical analysis confirmed that there was not a significant interaction between ENC Ability Scores at T1 and T2 ($p = .422$).	Statistical analysis confirmed that there was not a significant interaction between ENC Ability Scores at T1 and T2 ($p = .377$).	Statistical analysis confirmed that there was not a significant interaction between Mathematics CK Scores at T1 and T3 ($p = .317$).

Checks for Research Questions 3 and 4

Assumption and Method of Testing	T2 Receptive Language Total Raw Score on CELF-P2	T2 Teacher-Rated AtL Initiative Raw Score
1. Level of the Data	Ratio Level Data	It might be argued that data on Initiative questionnaires were ordinal level, as they were derived from Likert-scale judgements where the intervals between different values may not be equal. However, research indicates that parametric tests are robust enough to analyse Likert responses (Norman, 2010; Sullivan & Artino, 2013), particularly when derived from a published psychological measure and when compiled over several items (Coolican, 2014). They were therefore treated as at least interval level for the purpose of analysis.
2. Normality <ul style="list-style-type: none"> Visual analyses to explore Q-Q plots, identifying values deviating greatly from those expected in a normal distribution. Shapiro-Wilk tests to assess the null hypothesis that the data did not deviate significantly from normality. Normality can be assumed if $p > .05$ 	<ul style="list-style-type: none"> Q-Q plots indicated that data in both conditions broadly fitted what would be expected from a normal distribution, although one data point appeared to deviate slightly at the upper end of the distribution in both graphs. However, Shapiro Wilk tests were not significant for either condition. (Control $p = .398$; Intervention $p = .493$) so Normality was assumed. 	<ul style="list-style-type: none"> Q-Q plots indicated that a number of data points deviated away from the scores that would be expected in a normal distribution, particularly at the lower and upper ends of the distribution. Shapiro Wilk tests were not significant for the control condition, but they were for the experimental condition (Control $p = .119$; Intervention $p = .024$) Normality could not be assumed.

<i>Assumption and Method of Testing</i>	<i>T2 Receptive Language Total Raw Score on CELF-P2</i>	<i>T2 Teacher-Rated AtL Initiative Raw Score</i>
<i>3. Homogeneity of Variance</i> <ul style="list-style-type: none"> Levene's test to assess the null hypothesis that the variances from each sample do not differ significantly from each other. Homogeneity of variance can be assumed if $p > .05$ 	<p>The test was not statistically significant, $F(1, 39) = .13, p = .726$, and therefore homogeneity of variance was assumed.</p>	<p>Not conducted</p>
<i>4. Linear relationship between the DV and covariate</i> <ul style="list-style-type: none"> Scatterplots were constructed to visually analyse the relationship between both the DV and pre-test scores 	<p>Visual analysis of the scatterplot indicated an approximately linear relationship between the Receptive Language Total Raw Scores at T1 and T2, and that this assumption was not violated.</p>	<p>Not conducted</p>

<i>Assumption and Method of Testing</i>	<i>T2 Receptive Language Total Raw Score on CELF-P2</i>	<i>T2 Teacher-Rated AtL Initiative Raw Score</i>
<p><i>5. Homogeneity of regression across all experimental groups</i></p> <ul style="list-style-type: none"> Statistical testing to check that there was not a significant interaction between the treatment condition and covariate. Homogeneity of regression can be assumed if $p > .05$ 	<p>Statistical analysis confirmed that there was not a significant interaction between Receptive Language Total Raw Scores at T1 and T2 ($p = .308$).</p>	<p>Not conducted</p>

8.17 Thematic Analysis Step 1: Transcription Notation System

The notation system displayed here was used to transcribe the audio recordings of the semi-structured facilitator interviews. These notation features were adapted from Braun and Clarke (2013).

Notation Feature	Explanation of Use
Speaker's name:	The identity of the speaker of that turn in the conversation
((laughs))	Laughing during the turn of talk
((coughing))	Coughing during the turn of talk
((pause))	Pauses lasts 2-4 seconds
(.)	Short pause of less than a second
((long pause))	A pause lasting over 5 seconds
((in overlay))	Overlapping speech
((inaudible))	Speech/sounds that are inaudible
?	Rising intonation of a question
<u>word</u>	Emphasis placed on a particular word
“ “	Enclosing reported speech
[part of word]-	Indicates missing speech sounds, partially spoken word

8.18 Thematic Analysis Step 2: Example of Initial Data Coding from Anne's Interview Transcript

Data Source ²³	Data Extract	Codes
A6 Lines 56-66	<p>Interviewer: ((in overlay)) Yeah that's great thank you that's fantastic and so were you happy with the outcomes from the app?</p> <p>Anne: ((in overlay)) Yeah definitely I think toward the end erm some of the children got a bit fed up with it</p> <p>Interviewer: Ok</p> <p>Anne: ((in overlay)) Erm (.) you could tell you know they were just sat there and I was having to give them a bit more</p> <p>Interviewer: Mm</p> <p>Anne: ((in overlay)) Erm a few more prompts and er</p> <p>Interviewer: ((in overlay)) Some motivation</p> <p>Anne: Yeah definitely</p>	<p>Intervention lasted for too long as some children lost motivation</p> <p>Individual differences in children's motivation to use the intervention</p>
A7 Lines 67-80	<p>Interviewer: ...erm did you notice any differences in the way that (.) some of the children might have responded to the intervention?</p> <p>Anne: ((pause)) Yes a few a few of the children were very keen and they sort of saw it as a race</p> <p>Interviewer: Mm</p> <p>Anne: "I'm on I'm on this number now (.) you're only on that one" so obviously I spoke about how some will be</p>	<p>Peer competition motivated some children</p> <p>Individual differences in children's motivation to use the intervention</p>

²³ Each data extract was given a code e.g. A5, where A refers to the first initial of the pseudonym of the facilitator and 5 numerically labelled that section of the transcript. Line numbers identify the section of the full data transcript which the extract was taken from.

	<p>faster than others it's not a race we're just trying our best</p> <p>Interviewer: ((in overlay)) Yeah</p> <p>Anne: A lot of them got very very competitive with it which I suppose in a way is good</p> <p>Interviewer: ((in overlay)) Yeah</p> <p>Anne: Erm you know they wanted to do better than each other.</p>	
<p>A8</p> <p>Lines 82-94</p>	<p>Anne: (.) Erm some of the children were more (.) erm willing to ask for help if they were getting stuck</p> <p>Interviewer: Yeah</p> <p>Anne: "I need your help"</p> <p>Interviewer: ((in overlay)) Mm</p> <p>Anne: And they took the headphones off (.) and actually asked me for help which obviously was fine (.) erm a lot of the time with some of them it was just getting them to actually listen to what the lady was saying</p> <p>Interviewer: ((in overlay)) The instructions</p> <p>Anne: And I just sort of made them press the listen again button</p> <p>Interviewer: ((in overlay)) Mm</p> <p>Anne: And they were fine then so a lot of them just needed (.) the reassurance I suppose and just to listen again</p>	<p>Children supported by prompts to listen again.</p> <p>Differences in children's willingness to ask for help</p> <p>Children benefited from emotional support from the facilitator</p>

8.19 Thematic Analysis Step 2: List of Final Codes from Interviews

1. Intervention led to increased confidence with tablet technology
2. Intervention led to improved counting skills
3. Intervention led to improved number recognition
4. Intervention led to improved attention for some children
5. Motivation of some children declined over the course of the intervention
6. Some children motivated by peer competition which supported progress
7. Children supported by prompts to listen again
8. Differences in extent to which children asked for facilitator support
9. Children benefited from affective support from the facilitator
10. Individual differences in outcomes
11. Intervention developed confidence of some children
12. Children with initially lower attainment made more progress (contradiction)
13. Children benefited from additional cognitive support from facilitator
14. Some children required facilitator support to maintain attention
15. Children with lower attainment needed more facilitator support
16. Intervention was an assessment tool to inform instruction
17. Some children were upset by negative feedback/lack of progress
18. Children responded well to visual features of the app
19. Missed sessions affected progress
20. Individual differences in amount of facilitator support needed
21. Children worked collaboratively (contradiction)
22. Facilitator supported understanding of instructions/vocabulary
23. Software bug affected progress
24. Headphones did not always work
25. Intervention led to improvements in mathematics vocabulary of some children (contradiction)
26. Children with poorer language skills or EAL needed more support
27. Facilitator provided additional support with app quizzes
28. Children were motivated by positive feedback
29. Sessions were too long
30. Some children would have benefited from more physical and concrete experiences
31. Most (but not all) children enjoyed the intervention
32. Individual differences in children's motivation to use the intervention
33. Facilitator unsure about the impact of the intervention
34. Intervention did not have a noticeable impact on language skills
35. Some children with initially lower attainment made less progress
36. Children's attention skills affected progress
37. Children with lower language skills made less progress
38. Children found the technical features of the app easy to use
39. Charged tablets support ease of implementation
40. Activities of other children affected attention on-task
41. Facilitator's time to provide support affected progress
42. Variety of the activities/topics in the intervention was beneficial

- 43. Children would have benefited from more choice about when to use app
- 44. Children not generalising skills learnt during intervention
- 45. Intervention > improved shape recognition
- 46. Children's prior attitude to technology affected engagement
- 47. Children experiencing success developed confidence
- 48. Children with EAL benefited less from the intervention
- 49. Peer competition affected children's engagement with the content of the app (contradiction)
- 50. App content was developmentally appropriate
- 51. Progression of skills in app was not developmentally appropriate (contradiction)
- 52. Children were distracted by positive feedback from the app (contradiction)
- 53. Intervention could be better if aligned with whole class teaching
- 54. App did not provide sufficient scaffolding for some skills
- 55. Children were generalising skills in the environment (contradiction)
- 56. Children enjoyed experiencing success on the intervention
- 57. Intervention helped children to remain on-task with learning
- 58. Level of attention on-task depended on children's understanding of activity

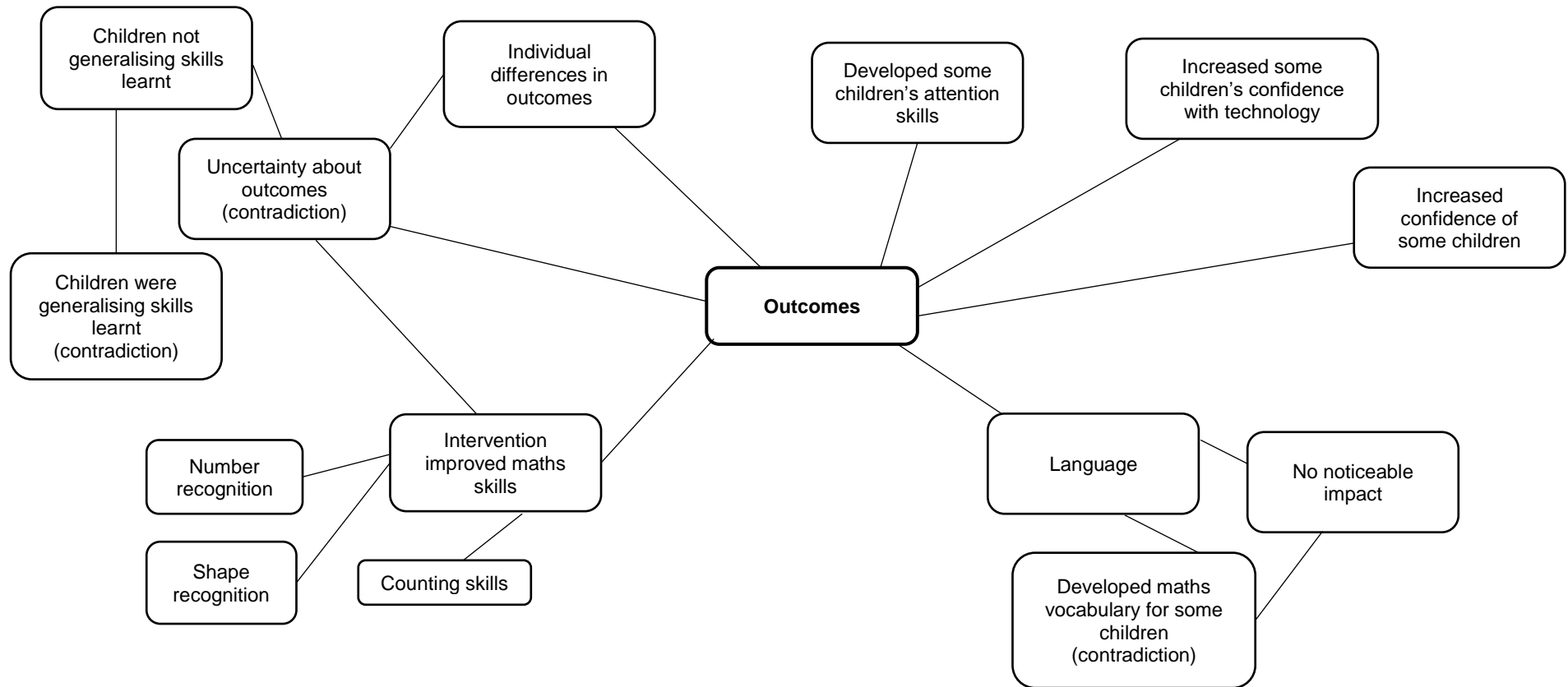
8.20 Thematic Analysis Step 2: List of Final Codes from Narrative Observations

1. Children navigated through the app easily
2. Child enjoyed receiving positive feedback
3. Child expressed confidence in their own ability after achieving success
4. Facilitator prompted child to maintain attention
5. Child sought reassurance from facilitator
6. Facilitator repeated the instruction
7. Facilitator provided affective support
8. Child expressed enjoyment in achieving success on app quiz
9. Peer competition
10. Child enjoyed activities on app
11. Child unsure how to complete activity
12. Facilitator supported child with app quiz
13. Child distracted by wider class environment
14. Child working independently through app activities
15. Child able to complete activity
16. Child did not pass app quiz
17. Facilitator not able to give sufficient support for app quiz
18. Facilitator supported understanding of app instructions
19. Facilitator not able to support children due to demands in the wider class environment
20. Child off-task
21. Child had a technical difficulty with app software
22. Facilitator gave pedagogical support
23. Child had technical difficulty with hardware
24. Facilitator prompted child not to rush
25. Facilitator praised child's progress
26. Child learnt from trial and error
27. Child not learning from trial and error strategy
28. Child not wearing headphones
29. Facilitator prompted child to listen again to app instructions
30. Facilitator did part of skill instead of child
31. Child off-task as did not understand activity
32. Child sought help with activity from facilitator
33. Facilitator gave additional support to child who had difficulties with counting/number recognition
34. Individual differences in level of facilitator support given during session
35. Peer support
36. Facilitator prompted child to focus on app content
37. Child used technical features of software appropriately
38. Facilitator removed child's headphones to listen to app instructions
39. Facilitator prompted child to return to activities for more practice before returning to quiz

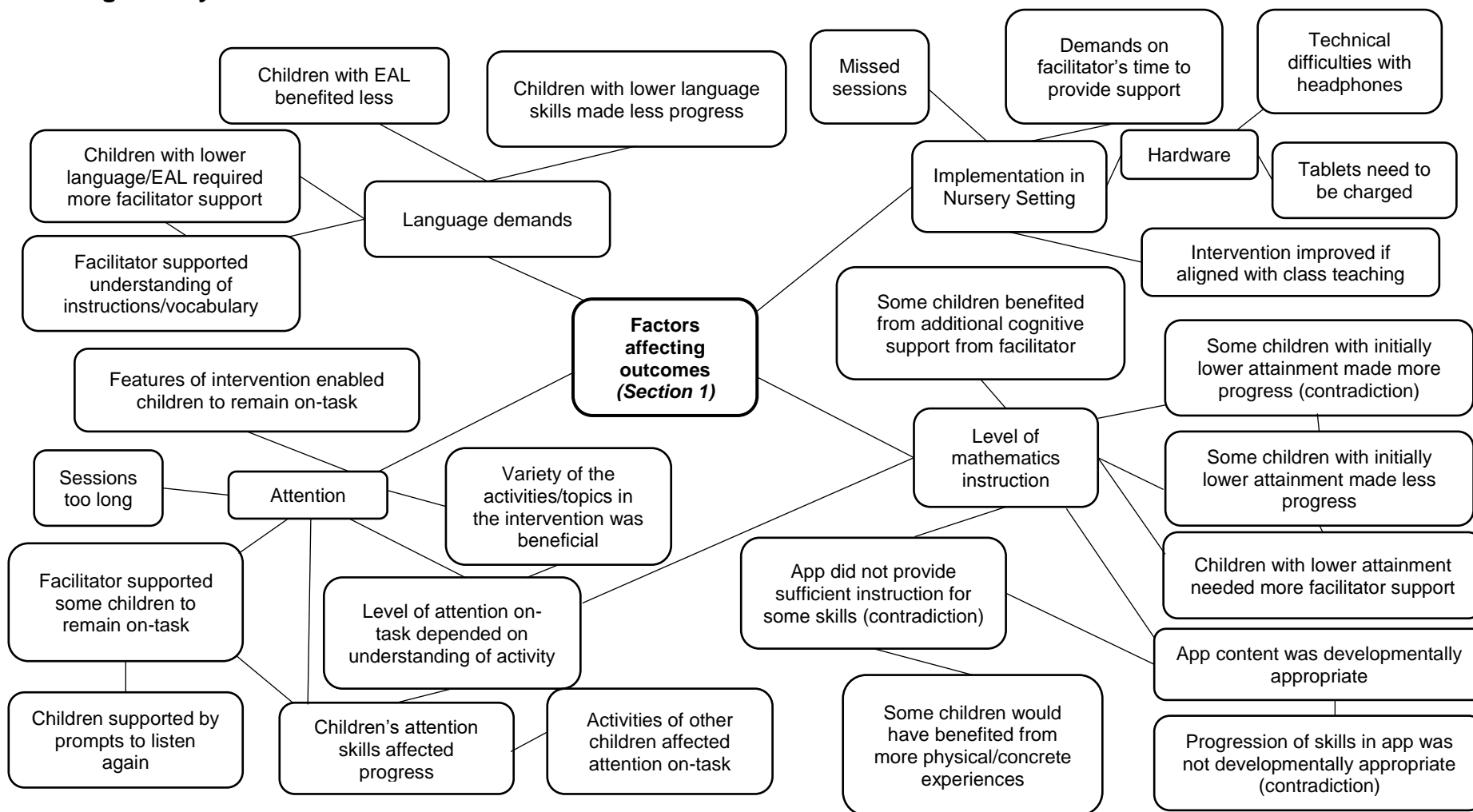
40. Facilitator prompted child to repeat app quiz
41. Facilitator prompted child to listen carefully to app instructions
42. Facilitator encouraged child to persevere
43. Child reacted to negative feedback
44. Child made a high number of errors on app quiz
45. Child sharing success with others
46. Facilitator gave technical support
47. Child expressed that they had reached a new level in the app
48. Child careful not to make mistakes
49. Child stopped quiz early to avoid failure after making a mistake
50. Child expressed enjoyment in progressing through software
51. Facilitator ensured child did not make a mistake on app quiz
52. Child looking at screen of another child
53. Child hesitant in responding on app
54. Software gave positive feedback incorrectly
55. Child did not pass app quiz on multiple occasions
56. Facilitator told child answer to app quiz

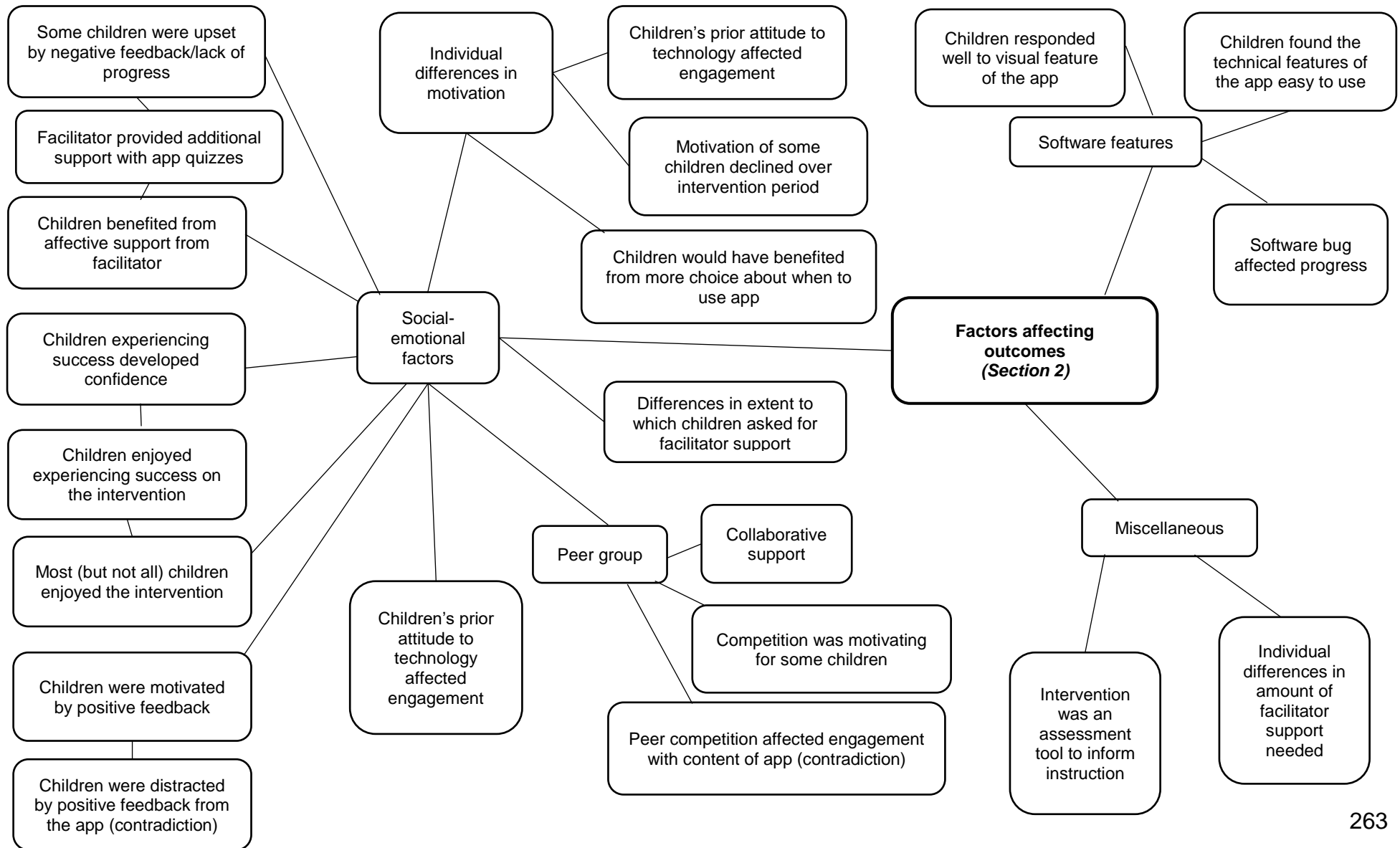
8.21 Thematic Analysis Step 3: Initial Thematic Maps of Interview Data

8.21.1 Research Question 5: What are facilitators' perceptions of the outcomes of the *onebillion* intervention for preschool children aged 3-4 years?



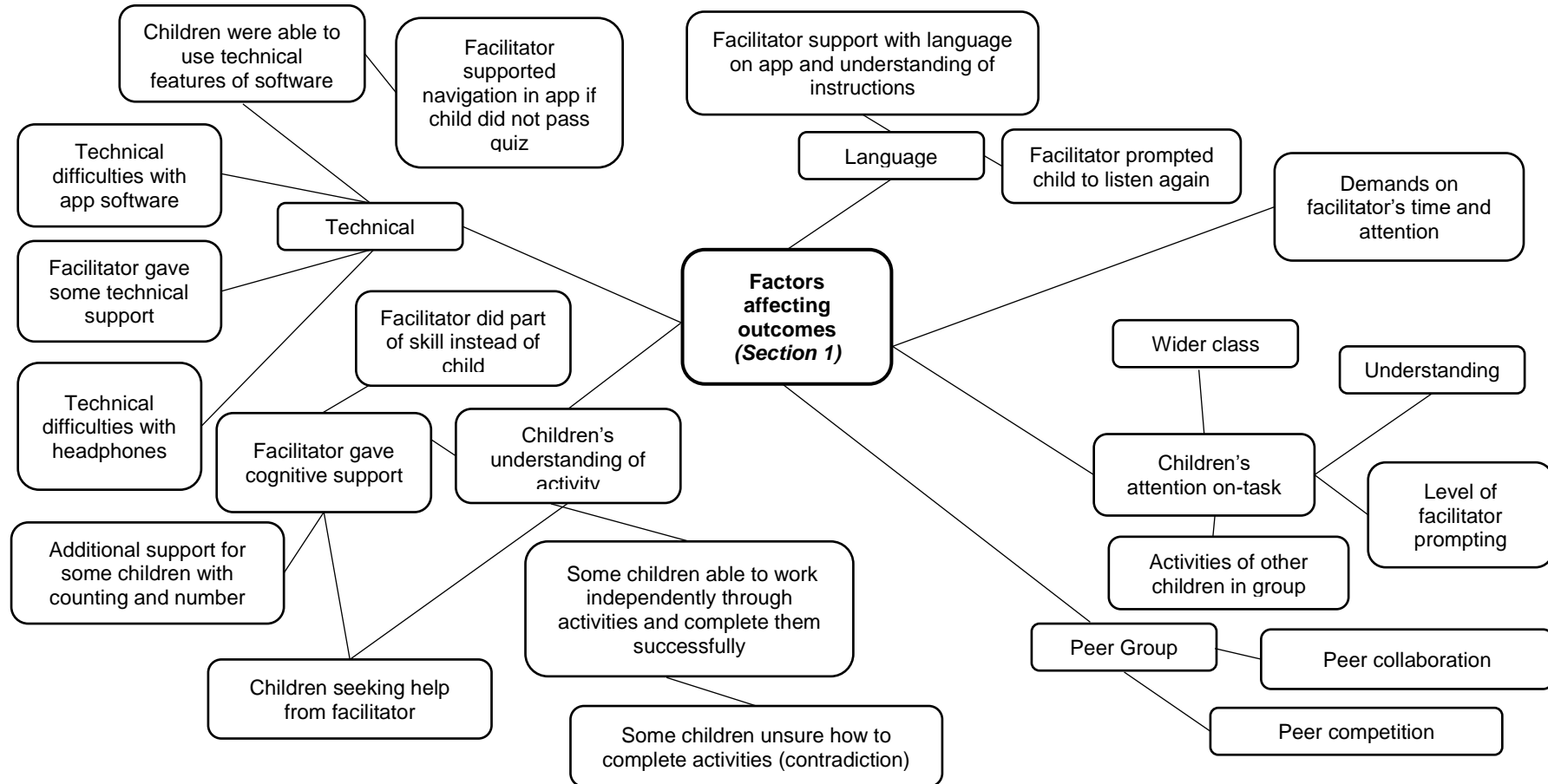
8.21.2 Research Question 6: What factors may affect the outcomes of the *onebillion* intervention for preschool children aged 3-4 years?

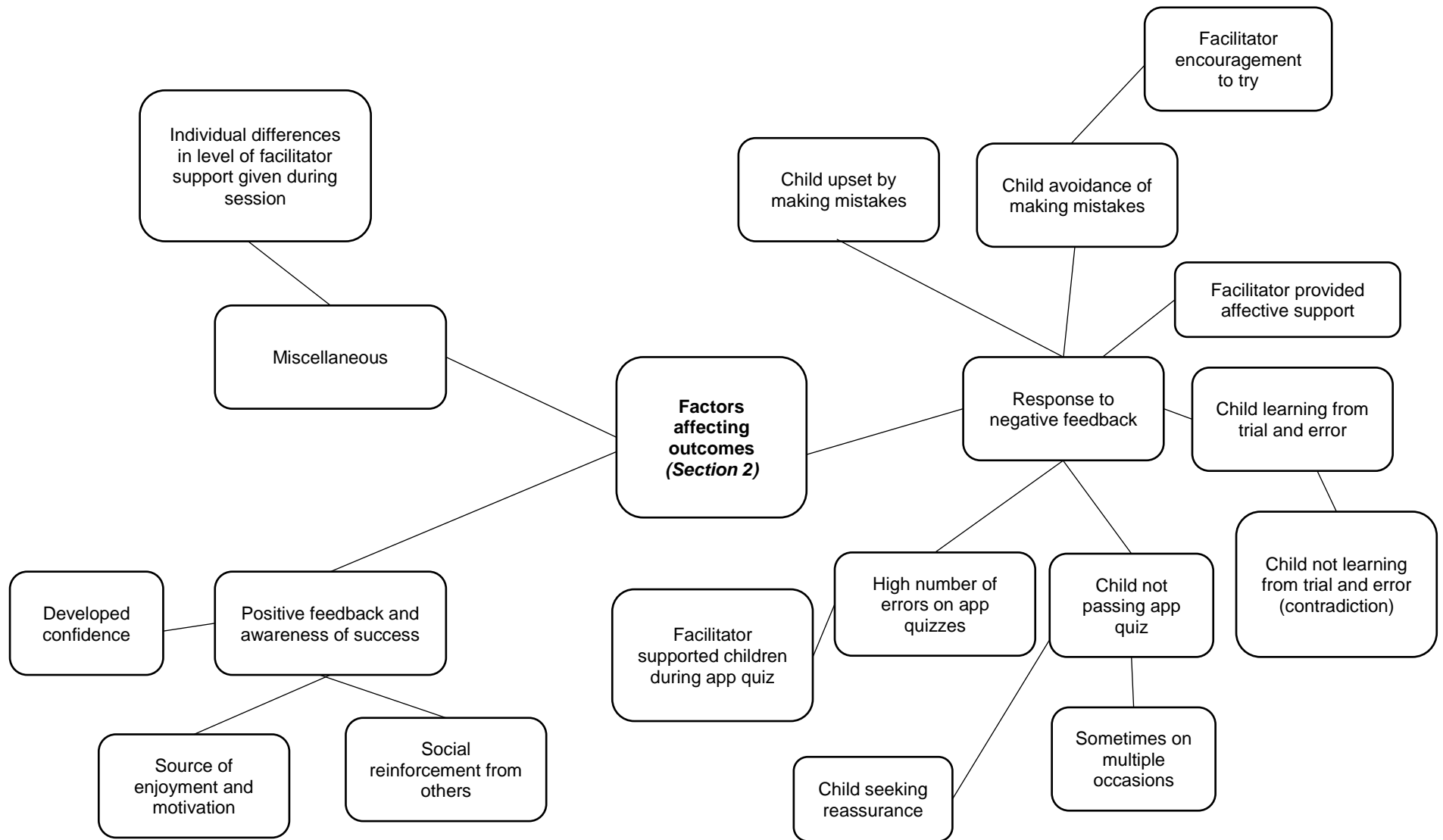




8.22 Thematic Analysis Step 3: Initial Thematic Map of Narrative Observational Data

8.22.1 Research Question 6: What factors may affect outcomes from the *onebillion* intervention for preschool children aged 3-4 years?





8.23 Thematic Analysis Step 4: An Example of Inter-Rater Reliability Checks for TA

Data Source ²⁴	Data Extract	Codes Green = Researcher Yellow = TEP 1 Pink = TEP 2
G5 Line 5-10	<p>T: Make sure you listen (went to support E doing a test)</p> <p>T: What did she say?</p> <p>E: (repeated back what virtual teacher said)</p> <p>T: What's one more than 17?</p> <p>E: (child chose correctly)</p> <p>T: Good girl</p>	<p>Facilitator prompted child to listen again to app instructions, or to listen more carefully</p> <p>Some children not passing app quizzes and/or requiring additional support</p> <p>Facilitator prompting (attention theme)</p> <p>Affective support</p> <p>Facilitator prompted child to listen again to app instructions, or to listen more carefully</p> <p>Some children not passing app quizzes and/or requiring additional support</p> <p>Affective support</p> <p>Facilitator prompted child to listen again to app instructions, or to listen more carefully</p> <p>Some children not passing app quizzes and/or requiring additional support</p>

²⁴ Each data extract was given a code e.g. G5, where G refers to the first initial of the pseudonym of the facilitator and 5 numerically labelled that section of the transcript. Line numbers identify the section of the full data transcript which the extract was taken from.

		Affective support from facilitator
G17 Line 17	T: (to D) Listen again, what did she say?	Facilitator prompted child to listen again to app instructions, or to listen more carefully Facilitator prompted child to listen again to app instructions, or to listen more carefully Facilitator prompted child to listen again to app instructions, or to listen more carefully
G34 Line 34	H: (cheered with arms when completing items correctly on the test)	Attitude dependent on awareness of success/positive feedback Attitude dependent on awareness of success/positive feedback Attitude dependent on awareness of success/positive feedback
G41 Line 41	F: Which is the lightest? (to G) Which would I be able to pick up really easily?	Facilitator supported with understanding of instructions/vocabulary Facilitator supported understanding of instructions/vocabulary Some children required additional cognitive support Facilitator supported with understanding of instructions/vocabulary

8.24 Comparison between Intervention Topics and BAS-III Early Number Concepts Assessment

This appendix supports comparison between the topics taught by the *onebillion* intervention and those assessed by the ENC sub-test of the BAS-III (Elliot & Smith, 2011).

<i>Maths Age 3-5 Topics</i>	BAS-III ENC Sub-test
Topic 1: Sorting and matching (by type, shape, colour and size)	Yes (size only)
Topic 2: Counting to 3 (counting, number recognition and formation)	Yes (not formation)
Topic 3: Lines and patterns (completing a simple pattern of pictures/sounds)	No
Topic 4: Counting 4 to 6 (counting, number recognition and formation)	Yes (not formation)
Topic 5: Where is it? (prepositional language)	No
Topic 6: Counting 7 to 10 (counting, number recognition and formation)	Yes (not formation)
Topic 7: Patterns and Shape (pattern recognition by shape/colour, 2D shape recognition)	No
Topic 8: Counting 1 to 10 (counting, identifying missing numbers, one more /one less)	Yes (counting and one more/one less only)
Topic 9: Comparing (language of comparison – size, quantity)	Yes
Topic 10: Adding and taking away (counting on, equation symbols)	Yes (addition only)

First five of the <i>Maths Age 4-6</i> Topics	BAS-III ENC Sub-test
Topic 1: Shape and Position (symmetry, turns, positional language)	No
Topic 2: Counting to 20 (counting, number recognition and formation)	Yes (not formation)
Topic 3: Sharing (sharing into sets and groups)	No
Topic 4: More Counting (ordinal numbers, counting in twos, odd/even)	Yes (ordinal numbers, counting in twos only)
Topic 5: Telling the time	No