Education
Endowment
Foundation

Early Years and Key Stage 1 Mathematics Teaching: Evidence Review

August 2020

Jeremy Hodgen (UCL Institute of Education)
Nancy Barclay (University of Brighton)
Colin Foster (Loughborough University)
Camilla Gilmore (Loughborough University)
Rachel Marks (University of Brighton)
Victoria Simms (Ulster University)

UCL Institute of Education
University of Brighton
Loughborough University
Ulster University

Please cite this report as follows: Hodgen, J., Barclay, N., Foster, C., Gilmore, C., Marks, R. and Simms, V. (2020). *Early Years and Key Stage 1 Mathematics Teaching: Evidence Review*. London: Education Endowment Foundation. The report is available from:

https://educationendowmentfoundation.org.uk/public/files/Early_Years_and_Key_Stage_1_Mathematics_Teaching_Evidence_Review.pdf

Acknowledgements

The authors would like to thank the peer reviewers, Dr Thérèse Dooley (Dublin City University), Professor Steve Higgins (Durham University) and Professor Lieven Verschaffel (KU Leuven); the guidance panel, Dr Sue Gifford (University of Roehampton), Dr Terezinha Nunes (University of Oxford), Mari Palmer (North Yorkshire Coast Research School), and Viv Lloyd (NCETM); and Amy Clark (EEF) for her advice and support.

Contents

Glossary of abbreviations

ACME	Advisory Committee on Mathematics Education
BES	Best Evidence Synthesis
CAI	Computer-assisted instruction
CI	Confidence interval
CK	Content knowledge
CPA	Concrete-Pictorial-Abstract
DI	Direct Instruction
DREME	Development and Research in Early Math Education
EEF	Education Endowment Foundation
ES	Effect size
EY	Early Years
G1	Grade 1 etc
K	Kindergarten
KS1	Key Stage 1
NCETM	National Centre for Excellence in Teaching Mathematics
NK	Not known
PCK	Pedagogical content knowledge
PD	Professional development
Pre-K	Pre-Kindergarten
QED	Quasi-Experimental Design
RCT	Randomised Controlled Trial
SD	Standard deviation
TA	Teaching assistant
WWC	What Works Clearinghouse

Introduction

This document presents a review of evidence commissioned by the Education Endowment Foundation to inform the writing of the guidance report *Improving mathematics in the Early Years and Key Stage One.*

The review aimed to synthesise the best available international evidence regarding teaching and learning of mathematics for children in Early Years and Key Stage 1 (i.e., between the ages of 3 and 7) and aimed to answer the following research question:

> *What is the evidence on the effectiveness of classroom-based interventions for improving mathematical learning of children in Early Years and Key Stage 1 settings?*

Over the past decade or so, there have been a number of 'best evidence' reviews that have surveyed and synthesised the research evidence on how young children learn mathematics in order to consider how teaching could be adapted to better support learning (e.g., Clements et al., 2013; Cross et al., 2009; Deans for Impact, 2019; Dooley et al., 2014). These reviews provided valuable context for our review and enabled us to triangulate our findings with the wider literature on mathematics learning. However, our review took a different approach by focusing principally on teaching. Specifically, we reviewed the experimental evidence about the efficacy of teaching interventions designed to improve children's learning in mathematics. We urge readers to view our review as a complement to the existing body of 'best evidence' and to consider the findings of this review in the context of this wider literature.

For the purposes of this review, a teaching intervention is defined as a change to existing classroom practice. This covers a broad range of interventions, from relatively 'small-scale' strategies, such as the use of manipulatives, to 'large-scale' programmes that are intended to cover the entire Early Years curriculum for a term or more. The critical characteristic is that the intervention is clearly described and could be implemented in Early Years settings (perhaps with some modification and in some cases with substantial costs). The interventions are grouped into 'strands', each addressing a specific topic.

For the purposes of informing the guidance, we were additionally asked to review the evidence in several areas that went beyond the strict classroom-based focus. First, we were asked to consider the evidence on children's progression in mathematics between the ages of 3 and 7. As a result, we developed three diagrams illustrating development in number, operations and geometry and spatial reasoning. Second, we were asked to examine the evidence about interventions addressing professional development for teachers and other Early Years educators (including professional knowledge), interventions to support the transitions from Early Years to Key Stage 1 and from Key Stage 1 to Key Stage 2, parent and family numeracy programmes and grouping by attainment. We report on all of these strands except grouping by attainment. We identified no evidence about grouping by attainment, either from the experimental literature or from the wider 'best evidence' literature. Hence, we cannot comment on the effectiveness of either grouping or setting by attainment, or alternatively mixed attainment practices, except to say that this does not appear to be an active question for early years researchers or educators.

Our aim was to focus primarily on causal evidence of impact from robust experimental or quasi-experimental studies. Given the rapid timescale for the review, our main focus was on the effects of different interventions on attainment rather than on attitudes or other non-cognitive outcomes. Using a systematic literature search strategy, we identified 115 relevant studies with their results reported in sufficient detail to be included a meta-analysis. However, there were nevertheless significant gaps. For example, there was very limited experimental evidence about the effectiveness of interventions that support either play and mathematical talk, both of which are shown by the wider literature to be important factors in children's mathematical learning (e.g., Clements et al., 2013). To address this, we supplemented our main dataset with 12 systematic and other 'best evidence' reviews.

The review built upon, and extends, three existing reviews carried out by members of the team and others:

- Simms, V., McKeaveney, C., Sloan, S., & Gilmore, C. (2019). Inter*ventions to improve mathematical achievement in primary school-aged children*. London: Nuffield Foundation.
- Hodgen, J., Foster, C., Marks, R., & Brown, M. (2018). *Evidence for Review of Mathematics Teaching: Improving Mathematics in Key Stages Two and Three: Evidence Review*. London: Education Endowment Foundation.
- Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide* (NCEE 2014-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Finally, much of the experimental evidence is from studies of interventions conducted in the United States and mainland Europe, rather than in the UK or England. In some cases, particularly for 'large-scale' programmes from the US, these interventions would require significant modification to the content and language in order to be used on a widespread basis in England. To address this, for each strand, we consider the relevance of the evidence base to Early Years and Key Stage 1 settings in England.

The structure of this document

We begin with a section about the development progression diagrams. This is followed by sections that outline our findings for each of the 18 interventions or strands, presented in 15 sections, broadly in order of the quality of the evidence base.[1] Finally, we present our methodology, which is supplemented by several appendices.

Terminology

In this document, we refer to 'educators', except where the evidence refers specifically to teachers, teaching assistants or other adults.

---

[1] Note that three pairs of related interventions (computer-assisted instruction, apps and technology tools, mathematical talk and the use of storybooks, and peer tutoring and cooperative learning) are presented together.

Guide to Reading the Review

Meta-analyses and effect sizes

As we have observed in the introduction, our approach was to focus primarily on reviewing causal evidence of impact from robust experimental or quasi-experimental studies. Our aim was (where possible) to carry out meta-analyses to estimate the effect (or impact) of the interventions identified.

Meta-analysis is a statistical procedure for combining data from multiple studies. If a collection of studies are similar enough, and each reports an effect size, the techniques of meta-analysis can be used to find an aggregated (or overall) effect size that indicates the best estimate of the underlying effect size for all of those studies.

In education, effect size (ES) is usually reported as Cohen's *d* or Hedges' *g*, which are measures of the difference between two groups in units determined by the standard deviation (the variation or spread) within the groups. An effect size of +1 means that the mean of the intervention group was 1 standard deviation higher than that of the control group. In practice, an effect size of 1 would be extremely large, and typical effect sizes of potential practical significance in education tend to be around the 0.1-0.5 range. Given our focus on experimental and quasi-experimental studies, we have reported effect sizes using Cohen's *d*.

Caution should be exercised in comparing effect sizes for different interventions which may not be truly comparable in any meaningful way. Judgment is always required in interpreting effect sizes, and it may be more useful to focus on the order of related effect sizes (higher or lower than some other effect size) rather than the precise values. It should be noted that effect sizes are likely to be larger in small, exploratory studies carried out by researchers than when used under normal circumstances in schools. Effect sizes may be artificially inflated when the tests used in studies are specifically designed to closely match the intervention, and also when studies are carried out on a restricted range of the normal school population, such as low attainers, for whom the spread (standard deviation) will be smaller.

In this review, in addition to reporting the effect size with a 95% confidence interval, we have categorised the effect sizes as small, medium or large, in line with previous work by some members of the team (Hodgen, Coe, Foster, Brown, Higgins, & Küchemann, 2020). This broadly follows Cohen's (1988) rule of thumb: effects of below d=0.05 as negligible, effects up to d=0.25 as small, effects of $0.25 < d < 0.75$ as moderate and effects of d=0.75 or greater as large. It was judged possible to carry out a meta-analysis for only seven of the 18 interventions that were identified. Of these, six of the effects were categorised as moderate and one as large.

See the Methodology section and the various more detailed appendices for further information on how the meta-analyses were carried out.

Quality (or strength) of the experimental evidence base

The quality (or strength) of evidence assessments were based on the GRADE system in medicine (Guyatt et al., 2008). This is an expert judgment-based approach that is informed, but not driven, by quantitative metrics (such as number of studies included). These judgements took account of several factors: the number of original studies, the methodological quality of the original studies, the consistency of results across studies, and any reporting bias, as well as additional evidence from systematic reviews and best evidence syntheses. Each member of the review team made independent judgments, which were then compared, aggregated and moderated.

Grades were on a scale from 0 (minimal) to 3 (strong). Whilst the approach was primarily judgment-based, we did operate thresholds for the strong, weak and minimal grades.

The experimental evidence base could be graded as ***strong*** only if we identified at least 20 experimental studies that met our inclusion criteria, two of which were conducted at scale. (See Apoendix 4.) For a strong grade, there needed to be, in our judgment, sufficient evidence from the remaining factors to support this grade. In the event, none of the strands was judged as having a strong evidence base, although three (computer-assisted instruction (CAI) and apps, explicit teaching, and individual / small-group tutoring by adults) were graded as having a moderate-to-strong experimental evidence base.

The experimental evidence could be graded ***weak*** if we identified at least one experimental study that met our inclusion criteria.

The experimental evidence was graded ***minimal*** if we identified no evidence from sufficiently robust experimental studies through our systematic searches of the literature, although readers should note that, for one strategy graded as minimal (play), the wider evidence base from systematic reviews was judged to be of moderate strength.

For further information, see Appendix 4.

Relevance of the experimental evidence base

We also used a similar judgement-based system to assess the relevance of the experimental evidence for Early Years and Key Stage 1 settings in England. These judgements took account of several factors: where and when the studies were carried out, how the interventions were defined and operationalised, any focus on particular topic areas, the age of children and phase of education, and the ease of implementation. Relevance is not independent of the quality of the body of evidence, so the overall relevance grading could not be more than one grade higher than the quality of evidence grading.

As we did for the quality (or strength) gradings, each member of the review team made independent judgments of relevance, which were then compared, aggregated and moderated.

For further information, see Appendix 4.

Structure

For each module, we give a *headline*, summarising the key points, together with key *definitions*, followed by a narrative account of the main *findings*. After noting any links to other strands, we summarise our judgements about the quality (or strength) and relevance of the *evidence base*.

Overview of review findings

The table below summarises the findings of our review. In addition to our findings about impact on attainment, quality of the experimental evidence base and relevance to Early Years and Key Stage 1 setting in England, we also summarise our judgments about the supporting theoretical and empirical evidence from systematic reviews and best evidence syntheses. This is particularly important for two interventions, play, and executive functions and metacognition.

Strand / intervention	**Aggregated effect size (or impact on attainment)**	**Quality (or strength) of experimental evidence base**	**Relevance of experimental evidence to Early Years and KS1 settings in England**	**Additional theoretical and empirical evidence**
Computer-assisted instruction (CAI) and apps	Moderate	Moderate-to-strong	Moderate-to-high	Moderate
Explicit teaching	Moderate	Moderate-to-strong	Moderate-to-high	Moderate
Individual and small-group tutoring by adults	Moderate	Moderate-to-strong	Moderate-to-high	Moderate
Manipulatives and representations	Moderate	Moderate	Moderate	Moderate
Whole-curriculum interventions	Moderate	Moderate	Moderate	Moderate-to-strong
Feedback and formative assessment	Moderate	Weak-to-moderate	Moderate	Moderate
Use of storybooks (reported with mathematical talk)	Large	Weak-to-moderate	Moderate	Weak-to-moderate
Movement and gesture	N/A	Weak	Weak	Weak
Cooperative learning	N/A	Weak	Weak	Weak
Play	N/A	Weak	Minimal	Moderate
Technology tools (reported with CAI)	N/A	Minimal	Minimal	Weak
Mathematical talk	N/A	Minimal	Minimal	Weak
Peer tutoring	N/A	Minimal	Minimal	None
Executive functions and metacognition	N/A	Minimal	Minimal	Moderate
Parent and family numeracy programmes	N/A	Minimal	Minimal	Weak
Problem-solving	N/A	Minimal	Minimal	Weak
Professional development and teacher (educator) knowledge	N/A	Minimal	Minimal	Weak
Transitions	N/A	Minimal	Minimal	Weak

Understanding early mathematical development

Early mathematical development is important for children's current achievement and also for their future learning and life success (Duncan et al., 2007; Watts, Duncan, Siegler, & Davis-Kean, 2014). Therefore, as educators, it is important to be aware of the typical development of mathematical skills and concepts in order to understand what may be appropriate for teaching children in the Early Years and Key Stage 1.

Mathematical development involves acquiring procedural skills, conceptual understanding and factual knowledge across a range of topic areas, including number, operations, and shape and space. It also involves forming connections between operations and concepts, such as understanding that addition is the inverse of subtraction. Children also need to develop reasoning skills, where they use mathematical ideas, structure or principles to justify or explain methods or solutions, or to extend their existing knowledge to new areas (Donovan & Bransford, 2005; Kirkpatrick, Swafford, & Findell, 2001). Mathematical development relies not only on specific mathematical knowledge and skills but also on higher-level thinking skills (executive functions), such as working memory (e.g., being able to hold information in your mind and manipulate it) and inhibition (e.g., ignoring distracting whole number information when dealing with fractions) (see Bull & Lee, 2014). The experiences that children have with mathematical materials and activities may also influence development (Elliott & Bachman, 2018). In addition, children's interest, enjoyment and attitudes towards mathematics also affect their learning (e.g., Dowker, Cheriton, Horton & Mark, 2019).

Therefore, because mathematics relies on specific knowledge and complex thinking skills and is also influenced by children's experiences and attitudes, development in this area can take extended periods of time and may be very taxing for young children. Children may develop mathematics skills at different rates and specific skills may emerge in different orders. Importantly, even if children appear to be engaging in mathematical activities (e.g., reciting the count sequence), they may not have a full grasp of the underlying concepts (e.g., understanding the meaning of the numbers in the count sequence). A particular challenge for children involves understanding that numbers are made up of other numbers (additive composition); for example, that 6 can be made up of 5 + 1, or 3 + 3, or 2 + 4. Experience with different types of countable objects may help children develop this understanding (Nunes & Bryant, 1996).

It is important to recognise that children will use different strategies to solve problems throughout development and this will be influenced by both their mathematical knowledge and their general thinking skills. There is strong evidence for this in the learning of addition, for example. When asked to complete a problem such as 2 + 3 = ?, children may begin by counting through all of the numbers in each set and then the combined set, perhaps using their fingers or countable objects (1, 2, … 1, 2, 3, … 1, 2, 3, 4, 5). With increased counting proficiency, children will begin to count on from the first number of one of the sets (2, … 3, 4, 5). With further understanding that order is irrelevant for addition (commutativity) and increased working memory capacity, children will identify and count on from the larger number (3,… 4, 5). Finally, after sufficient experience with a problem, children will retrieve the answer (5) (see Carpenter & Moser, 1984; Noel, Seron, & Trovarelli, 2003). This is not to say that any of the strategies used throughout development are incorrect; they are stepping stones towards becoming more efficient at completing problems, but they also place different demands on general higher-level thinking skills that develop over time.

We have summarised in three diagrams (Figures 1-3) what researchers currently know about the development of different key areas of early mathematics: number, operations, geometry and spatial thinking and measurement. The diagrams are spiral to convey that, whilst this hierarchy of steps is useful, development does not take place in clearly defined linear steps. Instead, skills and concepts develop in overlapping ways and children may develop some skills together or in a slightly different order. Moreover, developing a 'secure' grasp of these key mathematical ideas takes time. As a result, children's understanding may appear to differ in different settings or from one day to the next (Pirie & Kieren, 1994). The inner circle in each diagram indicates the general thinking skills or environmental influences that researchers have identified as being associated with development in this key area. The outer spiral highlights individual skills or concepts that develop over time. We have also provided specific examples of how our knowledge of development may influence what happens in the classroom, with teachers planning activities for learning, observing children's interaction with materials and then modelling or playing to support learning from these activities.

Further reading

In writing this section, we drew heavily on two recent syntheses that describe young children's mathematical development. For more information on the research underpinning the diagrams of children's development, please see these publications:

> Clements, D., Baroody, A. J., & Sarama, J. (2013). *Background research on early mathematics. Background Research for the National Governor's Association (NGA) Center Project on Early Mathematics*. [This outlines 'learning trajectories' for number and operations, and spatial reasoning, together with more explanation and illustrative examples.]
>
> Cross, C T, Woods, T A., & Schweingruber, H (Eds) [National Research Council] (2009). *Mathematics Learning in Early Childhood: Paths Toward Excellence and Equity*. Washington, DC: The National Academies Press. https://doi.org/10.17226/12519. [This provides detailed 'learning paths' for number, relations, and operations, and geometry and measurement, together with examples and further explanation.]

Additional evidence was drawn from:

> Deans for Impact (2019). *The Science of Early Learning*. Austin, TX: Deans for Impact.
>
> Gilmore, C., Gobel, S., & Inglis, M. (2018). *An Introduction to Mathematical Cognition*. London, UK: Routledge.

In addition, the Department for Education has recently published non-statutory guidance on progression in key mathematical concepts across primary, including previous experience required as the basis for learning in Year 1:

Department for Education / National Centre for Excellence in Teaching Mathematics. (2020). *Mathematics guidance: Key Stages 1 and 2. Non-statutory guidance for the national curriculum in England. Year 1*. London, UK: Department for Education.

Learning Progression Diagrams

Figure 1: Number

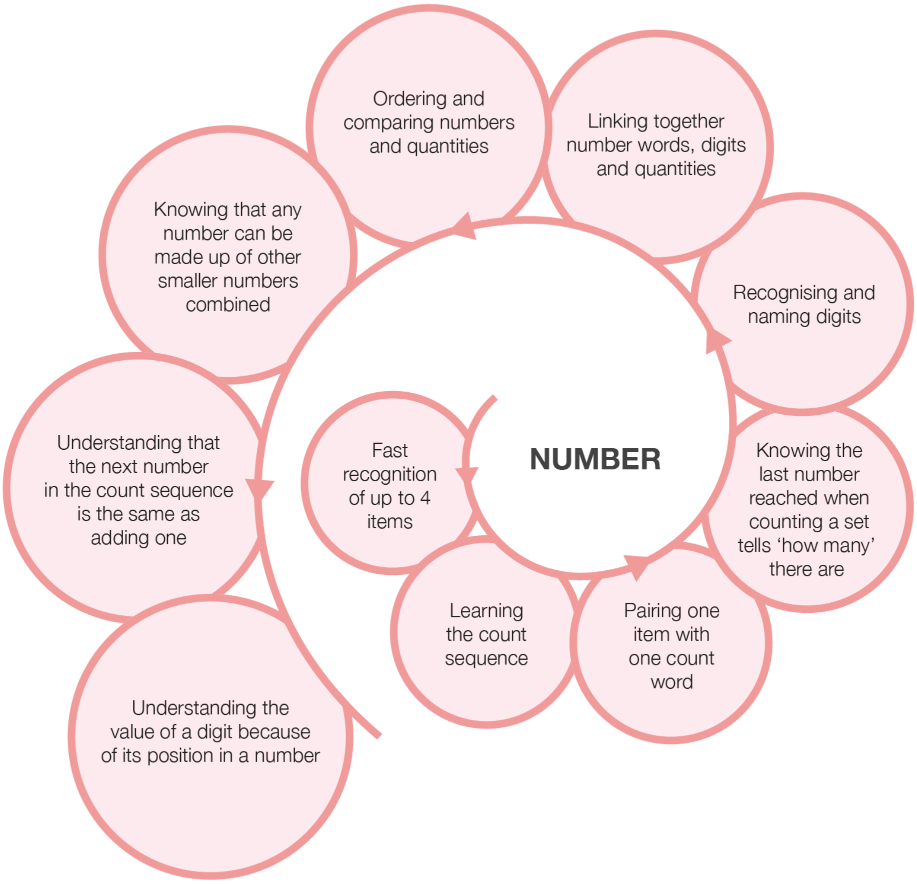


Figure 2: Operations

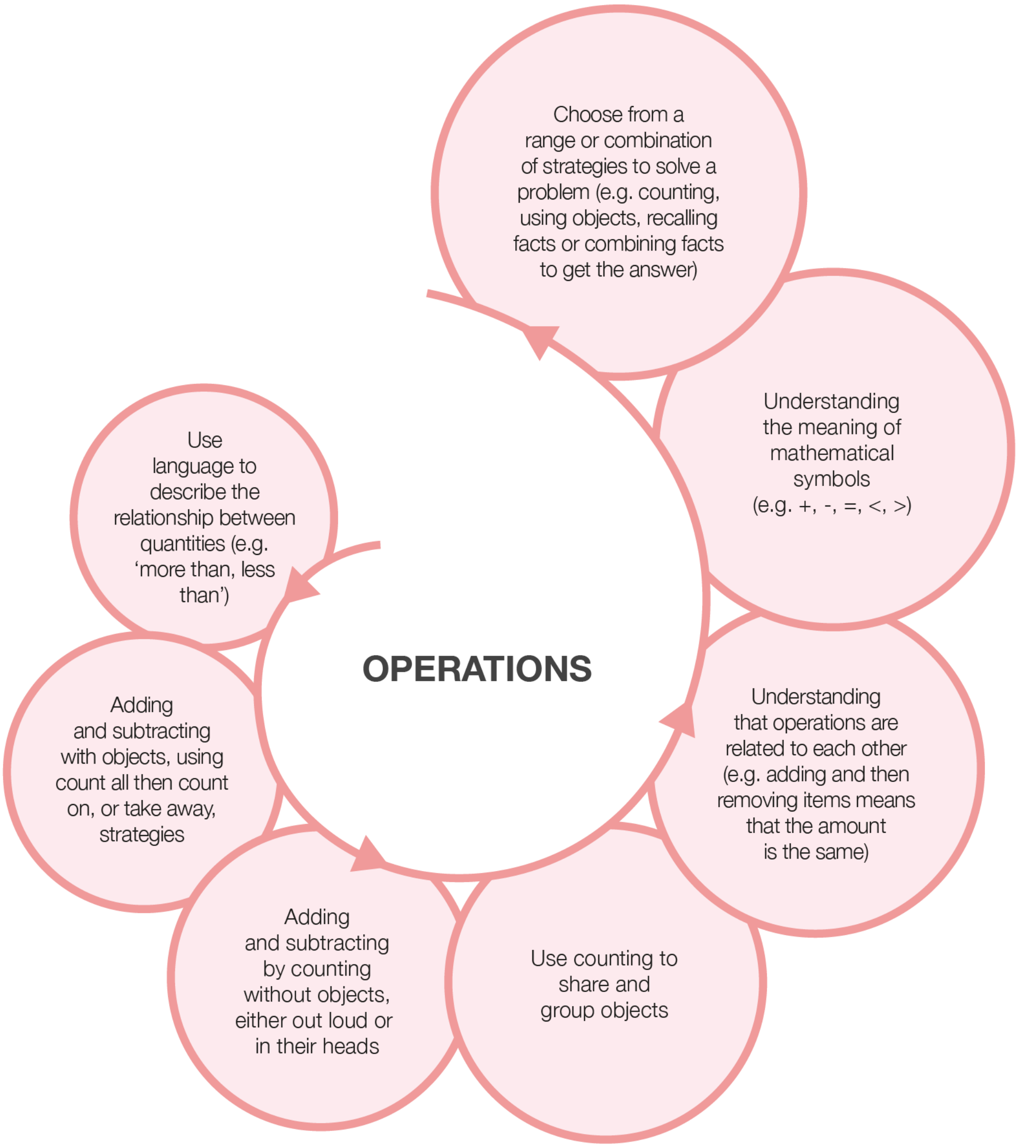
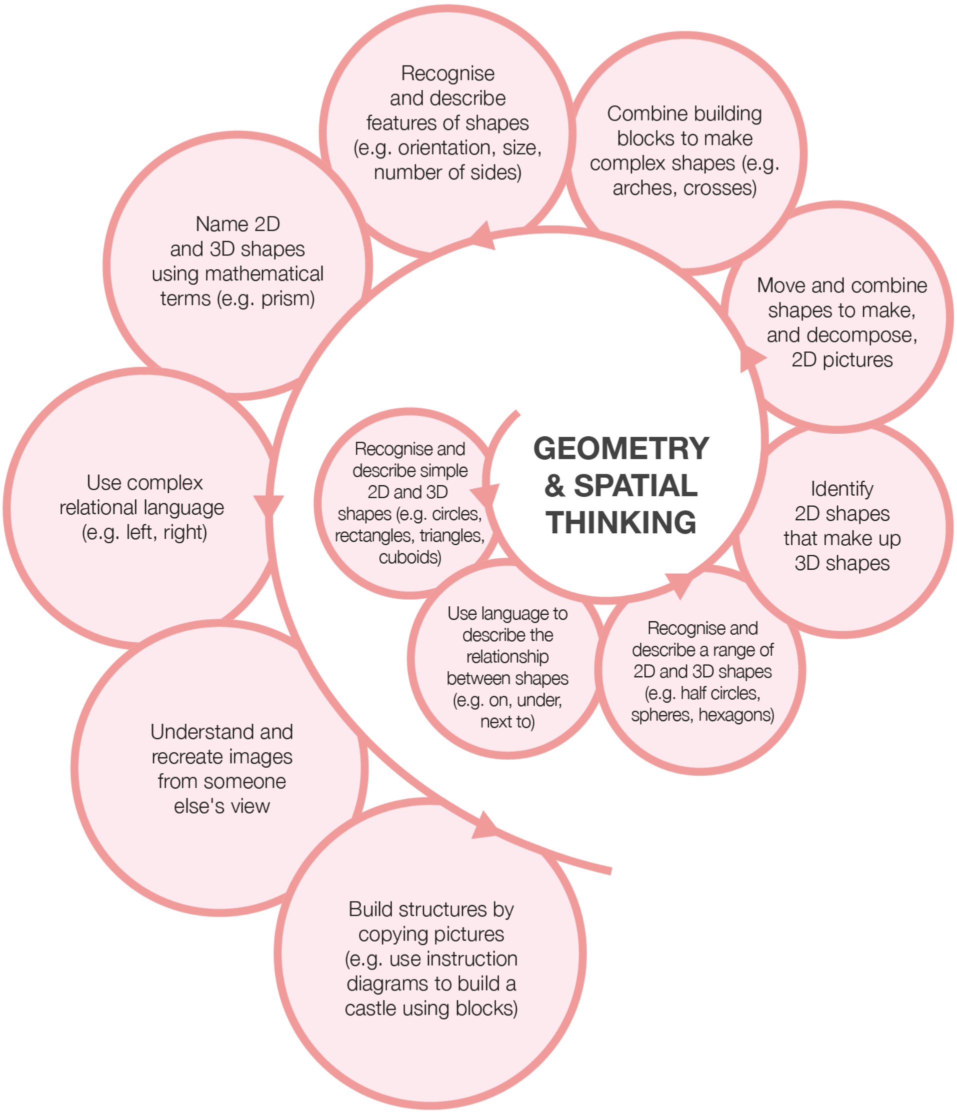


Figure 3: Geometry and Spatial Thinking



Computer-assisted instruction, apps and technology tools

There is a large body of evidence demonstrating that interventions delivered through Apps or Computer-Assisted Instruction (CAI) can have a positive effect on children's attainment in mathematics. However, much of the evidence relates to software that is not distributed in England or designed for the English mathematics curriculum. Hence, while such software has potential benefits, there is a need for research evaluating the effects of Apps and CAI specifically developed to align with the mathematics curriculum in England. The existing evidence base on CAI is relevant to both Early Years and Key Stage 1 settings. In contrast, there is very limited research examining the effect of using technology tools to support mathematics learning in either Early Years or Key Stage 1.

Definitions:

Computer-assisted instruction (CAI) and Apps: We use these terms interchangeably to refer to a broad range of computer- or tablet-based software designed to 'teach', or provide practice, for all or part of the mathematics curriculum. The software usually provides a controlled environment and often includes some form of corrective feedback to children. It may be set in the context of a specially designed game.

Virtual manipulatives refers to digital representations that attempt to mimic or model the movement of concrete manipulatives.

Technological tools refers to tools that can be used by children to explore and do mathematics, such as calculators, robots, programming software, dynamic geometry and digital technologies.

Findings:

Computer-assisted instruction (CAI) and Apps: We identified 37 studies, with 40 effects, that examined the effect of CAI and Apps and that met our inclusion criteria and had sufficient data to aggregate the reported effects in a meta-analysis. [This gave an overall moderate effect (d=0.42, 95% CI: 0.24, 0.59). All but three of the effects were positive and the interventions were in both Early years and Key Stage 1 relevant settings. This effect is larger than the small effect found for older pupils in an earlier review of mathematics interventions for Key Stage 2 and Key Stage 3 pupils (Hodgen et al., 2018).

There was a relatively high degree of heterogeneity across the effects that may reflect variation in the nature and scope of the studies and the different forms of CAI or Apps investigated. For example, some studies involved comprehensive interventions, in particular the *Building Blocks Software Suite*, which is designed to improve understanding and skills across the mathematics curriculum (Foster et al., 2016, 2018) and to facilitate child-initiated and open-ended activity. In contrast, others investigated much more focused interventions, often as part of an experiment investigating children's learning (e.g., Maertens et al., 2016, which is focused specifically on number lines).

Much of the evidence relates to software that is not available in England, either because the software has been developed in another educational system (often the US) or because the software has been produced for research purposes and is not commercially available. We identified only one App, *onebillion*, that has been evaluated in England. *Onebillion* addresses topics across the mathematics curriculum and has been the subject of a small-scale pilot (Outhwaite et al., 2019) and an independent efficacy trial (Nunes et al., 2019). As a result of the efficacy trial, *onebillion* is judged by the Education Endowment Foundation to have promise for mathematics teaching and learning.

Many of the CAI and App interventions involved the use of virtual representations, apparently intended to mimic concrete manipulatives, although it was not clear the extent to which children were afforded opportunities to 'manipulate' these virtual representations. There were no studies that directly compared the effects of virtual and concrete manipulatives.

Given the prevalence of experimental studies investigating the use of CAI or Apps, it is somewhat surprising that the Best Evidence Syntheses and expert reviews do not specifically address this type of software. However, these research syntheses all emphasise that software of any kind is likely to be more effective when combined with sound pedagogic practice.

Technological tools: Although the Best Evidence Syntheses and expert reviews support the use of technological tools, we identified just one experimental study relating to technological tools, specifically focused on computational thinking or programming. However, this study, Sung et al. (2017), which is also reported under the movement strand, did not investigate the effect of programming compared to a control, but rather compared different pedagogical approaches to providing teaching 'off the computer' to support the use of a programming language, *Scratch Junior*, in numeracy activities involving number lines. They found an effect for physical movement on mathematics outcomes and an effect for giving instructions about movement to a peer on computational thinking outcomes. We identified no experimental studies examining calculators, dynamic geometry or the broader range of technological tools.

Evidence from Best Evidence Syntheses (Clements et al., 2013; Cross et al., 2009) and expert reviews (Anthony & Walshaw, 2007, Dooley et al., 2014) provides additional support for the use of computer software, particularly highlighting the potential of virtual manipulatives and programming. However, all these syntheses also emphasise the importance of the educator's role and sound pedagogy in making best use of technological tools.

Cross et al. (2009) provide a useful checklist to help teachers select software with potential to aid mathematics learning, which is reproduced in Appendix 6.

Links to other strands:

Explicit teaching: As noted in the findings, in general, children need help from educators in order to develop sophisticated mathematical ideas.

Manipulatives and representations: As noted in the findings, many of the recent studies involved the use of virtual manipulatives.

Evidence base:

We judge the experimental evidence supporting the use of computer-assisted instruction and apps in general to be **moderate to strong**. We judge the experimental evidence supporting the use of technological tools to be **minimal**.

CAI and Apps

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	3	40, 4 at scale.
Methodological quality of the original studies	2.5	Mixed, but some good. Some independently evaluated studies.
Consistency of results	2	Heterogeneity. CAI / Apps covers a wide variety of approaches
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	2	Supports findings from the original studies
Overall Quality of Evidence judgment	**2.5**	**Moderate-to-strong**

Technological tools

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	1	1 study, insufficient information to aggregate
Methodological quality of the original studies	1	Small scale
Consistency of results	N/A	Too few studies to make judgment
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	1	No specific findings
Overall Quality of Evidence judgment	**0**	**Minimal**

Relevance:

The evidence on CAI and Apps is judged to be of **moderate** relevance to Early Years and Key Stage 1 mathematics teaching in England. There is insufficient evidence on technological tools to make a useful judgment on relevance.

CAI and Apps

Threat to relevance	Grade	Notes
Where and when the studies were carried out	2	Three studies (2 Apps) carried out in England. The remaining evidence is from a range of countries in America, Europe and the Far East.
How the interventions were defined and operationalised	2	Only 2 of the Apps evaluated are commercially available in England and many are designed for research purposes rather than for use in ordinary classrooms. Wide variety of different approaches.
Any focus on particular topic areas	3	Studies involved both number and/or calculation and geometry / spatial reasoning.
Age of children /phase of education	3	The studies were carried out across the age range and in contexts that have relevance to both Early Years and Key Stage 1.
Ease of implementation	2	Implementation requires appropriate infrastructure/ technology to be available
Overall relevance judgment	**2.5**	**Moderate to high**

Technological tools

Threat to relevance	Grade	Notes
Where and when the studies were carried out	N/A	Too few studies to make judgment
How the interventions were defined and operationalised	N/A	Too few studies to make judgment
Any focus on particular topic areas	N/A	Too few studies to make judgment
Age of children /phase of education	N/A	Too few studies to make judgment
Ease of implementation	N/A	Too few studies to make judgment
Overall relevance judgment	**0**	**Minimal**

Explicit teaching

There is a substantial body of evidence demonstrating that explicit, educator-led teaching has a positive effect on children's attainment, although structured teaching should be balanced with opportunities for children to engage in both free and structured (or guided) play involving mathematical resources and ideas. The evidence relates to a variety of educator-initiated interventions, ranging from direct instruction, which refers to tightly specified interventions often involving scripted lessons, to interventions that allow more flexibility. However, the evidence is largely based on interventions developed by expert teams for use in the US which involve considerable guidance and professional development for teachers and other adults delivering the intervention. In addition, much of the evidence concerns interventions with children assessed to be at risk of low attainment in mathematics or of mathematical learning difficulties. Although there is more evidence relating to interventions in the Early Years, the research on structured teaching is judged relevant in general to both Early Years and Key Stage 1. However, there is a need for more research and development examining how best to support the use of structured teaching in Early Years and Key Stage 1 contexts in England.

Definitions:

We use the term *explicit teaching* to refer to formal educator-directed approaches in which educators explicitly support children to develop specific mathematical ideas and skills. This covers a broad range of approaches that may be referred to as *direct instruction*, *explicit instruction* or simply as *structured teaching*.

Direct instruction is a term used in the US to refer to interventions that are often wholly or partially scripted, and which involve corrective feedback and structured practice, and which are taught on the basis of assessments that children have 'mastered' the necessary prerequisite knowledge. Direct instruction covers a wide range of approaches. At one extreme, *Direct Instruction* (capitalised) refers to a particular highly structured approach developed by Siegfried Engelmann and which has been evaluated in several relevant studies (Stockhard et al., 2017). In Early Years and Key Stage 1 settings, Direct Instruction follows a similar structure to later years, although teaching is normally in small rather than large (or whole-class-sized) groups and sessions may be shorter (e.g., McKenzie, Marchand-Martella, Moore & Martella, 2004).

Explicit instruction is a looser and broader term, also used in the US, that includes interventions involving high instructional guidance, where educators draw specific attention to mathematical concepts (for example, when using manipulatives, Carbonneau et al., 2013).

Structured teaching is used to refer to any planned, educator-led intentional teaching directed at a clear and specific learning goal (Cross et al., 2009). Educator-led does not imply a 'formal' classroom layout for teaching; often structured teaching may take place in 'informal' small groups. The key characteristics are that the teaching is planned and intentional, not that it takes place in a traditional classroom arrangement.

Findings:

We identified 26 studies, with 30 effects, that examined the effect of explicit teaching and met our inclusion criteria. There was sufficient data in the papers to aggregate 25 of these effects in a meta-analysis. This gave an overall moderate effect (d=0.66, 95% CI: 0.45, 0.87), although there was a relatively high degree of heterogeneity (I^2=93%) across the effects that suggests some variation in the effects of different interventions. All but one of the effects were positive and the interventions were in both Early years and Key Stage 1 relevant settings, although there is more evidence relating to interventions in the Early Years. One limitation is that most studies involved children screened as ‘at risk’ of low attainment or of mathematical learning difficulties and, therefore, it is difficult to generalise these findings to general classroom practice.

Evidence from several best evidence syntheses (Cross et al., 2009; Deans for Impact, 2019; Clements et al., 2013; Frye et al., 2013) support this finding, noting that it is unlikely that children will develop abstract or more sophisticated mathematical ideas without some structured teaching. Two expert reviews (Anthony & Walshaw, 2007; Dooley et al., 2014) add further weight to this evidence, although both place particular emphasis on the ‘planned’ elements of ‘structured’ teaching. However, more broadly, drawing on whole-curriculum approaches such as *Building Blocks* (Clements et al., 2011), these syntheses and reviews all strongly emphasise that free and structured play involving mathematics is important, not simply because this provides learning opportunities for children, but also because this provides opportunities to observe and assess children’s informal mathematical activity, as well as for engaging children in mathematics.

The evidence relates to a variety of educator-initiated interventions reflecting a range of different approaches and the bulk of the studies aggregated used some form of direct instruction, although this may be because such interventions are more clearly operationalised or manualised and thus are more amenable to investigation using experimental methods. Five studies involved Engelman’s Direct Instruction, but all were conducted prior to 2012. There is also a substantial body of evidence relating to structured teaching within wider interventions that place equal, or more, emphasis on educator judgment and child-initiated starting points for teaching (e.g., *Building Blocks*). We note also that Nelson & McMaster’s (2019) meta-analysis examining Early Years mathematics interventions did not identify an effect for direct instruction over and above the positive effect they identified overall for all the interventions considered.

The bulk of the interventions examined in these studies were designed and conducted in the US and relate to programmes that are not commercially available in England. In any case, these interventions, although largely well-designed, would not be appropriate for widespread use in settings in England without some adaptation to match the curriculum, UK English and the particular features and pedagogic traditions of Early Years and Key Stage 1 contexts in England. In addition, these interventions all involve both considerable guidance for educators (e.g., scripted lessons in Direct Instruction, Stockhard & Engelmann, 2010; or well-developed learning trajectories in *Building Blocks*) as well as substantial professional development support (e.g., regular monthly coaching in *Building Blocks*, Clements & Sarama, 2008). Hence, there is a need to develop, and evaluate, interventions designed specifically for the English context, examining both the effectiveness and the implementation of different forms of structured teaching, both in standalone forms and in the context of wider whole-curriculum approaches (as in *Building Blocks*).

Links to other strands:

Play: As noted above, the evidence from Best Evidence Syntheses indicates that structured teaching should be balanced with opportunities for play, which can include a diverse range of manipulatives (e.g., Deans for Impact, 2019) that provide opportunities for learning as well as for educators to observe, assess and thus plan appropriate structured teaching.

Feedback and formative assessment: Feedback and assessment are a key aspect of most explicit teaching interventions.

Evidence base:

We judge the experimental evidence supporting the use of explicit teaching in general to be **moderate to strong**.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	3	30, 2 at scale, many studies (41) excluded from meta, because insufficient information available to aggregate
Methodological quality of the original studies	2.5	Mixed, some good
Consistency of results	1	Some problems with definition, e.g., Building Blocks, a play-based curriculum, & Ramani Linear Game, both aggregated within some meta-analyses.
Reporting bias	1	Many interventions evaluated by developers
Evidence from systematics reviews and best evidence syntheses	**2**	Direct instruction somewhat contested, but support for explicit / planned teaching.
Overall Quality of Evidence judgment	**2.5**	**Moderate-to-strong**

Relevance:

The relevance of the evidence is judged to be **moderate-to-high.**

Threat to relevance	Grade	Notes
Where and when the studies were carried out	2	None of the studies in our dataset was carried out in England and almost all the evidence is from the US.
How the interventions were defined and operationalised	2	The studies combine a range of different approaches to structured teaching, many of which are manualised (but not for English contexts).
Any focus on particular topic areas	3	Studies involved both number and/or calculation and geometry / spatial reasoning. Some critiques of di argue that there is too much focus on procedures, but this is disputed (by, e.g., Gersten).
Age of children /phase of education	3	The studies were carried out across the age range and in contexts that have relevance to both Early Years and Key Stage 1.
Ease of implementation	2	Generally implementation is well-described or manualised. Requires PD in general.
Overall relevance judgment	**2.5**	**Moderate-to-high**

Individual and small-group tutoring by adults

There is a large body of evidence demonstrating that tutoring (or numeracy support) programmes delivered by teaching assistants, or by teachers, can have a positive effect on attainment for low-attaining children. However, this is likely to be the case only where the support is delivered through structured interventions that have been designed to address specific weaknesses in numeracy. Without this, evidence indicates that TA support has little, or even negative, effects on the attainment of low-attaining children. Almost all of these effective tutoring programmes have been developed by expert teams that have been informed by research on children's mathematical development and involve regular sessions over an extended period equivalent to a term or more. For these structured interventions, delivery by teaching assistants appears to be as effective as delivery by teachers. However, most of these interventions involve considerable guidance and professional development or coaching for tutors. The bulk of the evidence concerns interventions developed and evaluated in the US, although two of the programmes are available in England: Numbers Count and Mathematics Recovery. The evidence is relevant to both Early Years and Key Stage 1 contexts.

Definitions:

In this strand, we refer to face-to-face *tutoring* by educators in one-to-one, paired or small group settings targeted at low-attaining children. In Early Years settings and primary schools in England, this is often referred as either intervention or support for children who struggle with mathematics rather than tutoring and tutoring is often provided by teaching assistants. The experimental evidence relates to the efficacy of *structured programmes* that outline the support and/or teaching to be provided over a number of sessions.

Findings:

Tutoring by adults, particularly Teaching Assistants (TAs), currently plays a major role in education in England, especially in the support of low-attaining children using small-group or one-to-one tuition (Warhurst et al., 2013). Although there is a large body of evidence showing that, where educators use carefully designed and structured interventions, this can have a positive impact on attainment in numeracy for children at risk of low attainment, much of the support provided by TAs appears to be relatively unstructured (Sharples et al., 2015). There is a great deal of evidence indicating that such unstructured support has no, or even a negative, effect on learning (e.g., Blatchford et al., 2009). On the other hand, we found evidence that support, delivered through structured tutoring programmes, can have a positive impact on children's attainment.

We identified 15 studies, with 18 effects, that examined the effect of *tutoring by adults* in structured programmes that met our inclusion criteria and which reported sufficient data to aggregate the effects in a meta-analysis. This gave an overall moderate effect (d=0.50, 95% CI: 0.37, 0.64), although there was a high degree of heterogeneity across the effects that suggests some variation in the effects of the different programmes. The interventions were in both Early Years and Key Stage 1 relevant settings. All but one of the studies was conducted with children judged, or formally screened, to be at risk of low attainment in mathematics. This finding is supported by two meta-analyses conducted with primary age children:

Dietrichson et al.'s (2017) review of numeracy and literacy interventions for disadvantaged children, and Pellegrini et al.'s (2020) best evidence synthesis of mathematics interventions lasting 12 weeks or more. In addition, Cross et al.'s (2009) expert-judgment based review emphasises the importance of early intensive intervention to support children at risk of low attainment.

These interventions include *Number Rockets* (Rolfhus et al., 2013), *ROOTS* (Doabler et al., 2016) and *Galaxy Maths* (Fuchs et al., 2013). These programmes have been developed for use in the US by academic teams and have been informed by research both on children's development and on effective pedagogies. These programmes are extensive in nature; they cover a substantial element of the mathematics (or number and calculation) curriculum and involve regular structured tutoring sessions several times a week over an extended period equivalent to a term or more. Additionally they provide specific pedagogic guidance together with considerable professional development and/or instructional coaching (see, e.g., Kraft et al., 2018). *ROOTS*, for example, which has been extensively evaluated and is the focus of six studies, consists of 50 lessons, each of 20 minutes, delivered 5 days a week over 10 weeks, and tutors receive up to 4 coaching visits in addition to 2 days of professional development.

However, some caution should be exercised about the duration of tutoring interventions, and there is some evidence to suggest that time-limited tutoring support may be more effective. For example, Dietrichson et al.'s (2017) meta-analysis across primary, which identified tutoring as having a positive effect based on a large number of studies, found that longer interventions tended to be less effective. There is also evidence that the effects of interventions fade over time. For example, Smith et al.'s (2013) evaluation of *Mathematics Recovery* found that gains at the end of the programme were not maintained a year after the intervention. To address this problem, Clements et al. (2013) suggest that it is important for subsequent teaching to explicitly build on, and be consistent with, earlier teaching.

Pellegrini et al. (2020) found that, of the structured tutoring programmes they reviewed, all of which involved professional development for the tutors, those delivered by TAs were as effective as those delivered by qualified teachers. Pellegrini et al. also found no difference in the effect on attainment of interventions delivered individually and in small groups, although Wang et al. (2018) found a small difference in the effects.

Only one of the included studies was conducted in England: a large-scale randomised controlled trial of *Numbers Count*, conducted by an independent team of evaluators (Torgerson et al., 2013). This intervention is available in England, as is *Mathematics Recovery*, an intervention delivered by specially trained teachers, originally developed in Australia and independently evaluated at scale in the US (Smith et al., 2013). Additionally, *Mathematics Recovery* informed the development of *Catch Up ™Numeracy*, an intervention for pupils aged 6 to 14 delivered by teaching assistants in England, and which has been evaluated with older children, although with mixed results: an efficacy trial by Rutt et al. (2015) indicated a small but significant effect on attainment compared to a no-support control, whereas Hodgen et al.'s (2019) effectiveness trial found no effect when compared to a control group receiving matched time tutoring support.

Links to other strands:

Whole-curriculum interventions: Several whole-curriculum interventions, such as *Building Blocks* (Clements et al, 2011), involve small-group teaching as one element of the intervention. Several meta-analyses do not distinguish between tutoring interventions delivered individually or in small groups and whole-class interventions (e.g., Nelson & McMaster, 2018; Wang et al., 2016).

Explicit teaching: Several tutoring interventions, such as *ROOTS* (Doabler et al., 2016), involve direct or explicit instruction.

Evidence base:

We judge the experimental evidence supporting the use of tutoring for low-attaining children to be **moderate to strong**.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	3	18 effects from 13 studies; 3 studies at scale
Methodological quality of the original studies	2.5	Several well-constructed studies, although many programmes evaluated by developers
Consistency of results	1	Heterogeneity
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	2	Tutoring in small groups supported by BES, but little mention of para-professionals / TAs
Overall Quality of Evidence judgment	**2.5**	**Moderate-to-strong**

Relevance:

The relevance of the evidence is judged to be **moderate-to-strong.**

Threat to relevance	Grade	Notes
Where and when the studies were carried out	2	Aside from 2 in England, the identified studies were conducted in the US.
How the interventions were defined and operationalised	3	Mostly well-defined & manualised interventions.
Any focus on particular topic areas	2	Mostly number
Age of children /phase of education	3	The studies were carried out across the age range and in contexts that have relevance to both Early Years and Key Stage 1.
Ease of implementation	2	Generally implementation is well-described or manualised (although US-focused). Requires PD in general. Many use instructional coaching.
Overall relevance judgment	**2.5**	**Moderate-to-high**

Manipulatives and representations

Concrete manipulatives and representations are a powerful way of enabling young children to engage with mathematical ideas, provided that teachers and other adults help children to understand the links between the manipulatives or representations and the mathematical ideas that they represent through discussion and explicit teaching. There is consistent evidence that supports the use of physical manipulatives, and the evidence supports a variety of manipulatives, including linear board games, building blocks, counting aids and number lines. Linear board games, such as Snakes and Ladders, appear to be particularly beneficial for children from disadvantaged backgrounds or who struggle with numeracy. Children benefit from actually moving and interacting with manipulatives to understand mathematical ideas. As children's understanding of key mathematical ideas develops, educators should encourage children to use pictures, symbols and other more abstract diagrams to represent and communicate these ideas and concepts. Educators should show children different representations of number and make connections between them. Fingers provide a particularly valuable tool for supporting the understanding of counting, addition and subtraction. As with any intervention, educators need to consider carefully what manipulatives and representations to use and for how long in order to enable children to develop increasingly sophisticated mathematical ideas. The evidence relating to manipulatives and representations is consistent across, and relevant to, both Early Years and Key Stage 1.

Definitions:

Concrete manipulatives include counting aids (such as counters, unifix cubes or other objects, physical number lines, building blocks and board games), which can be 'manipulated' by children or adults, and may be physical or virtual (on a computer).

Representations include informal drawings (including drawings of manipulatives) as well as mathematical symbols (such as canonical dice patterns) and more formal mathematical diagrams (such as grids, drawn number lines and graphs). For the purposes of this review, we exclude gesture, which is included in the movement strand.

Findings:

We identified 19 effects from 17 studies that examined the effect of using manipulatives and/or representations and met our inclusion criteria with sufficient data to aggregate in a meta-analysis. The meta-analysis showed an overall moderate effect (d=0.34, 95% CI: 0.07, 0.60). There was a relatively high degree of heterogeneity across the effects that suggests some variation in the effects of different approaches and interventions. All but two of the effects were positive. Evidence from several best evidence syntheses (e.g., Cross et al., 2009; Deans for Impact, 2019; Clements et al., 2013) also support this finding.

Whilst evidence from best evidence syntheses is consistent in indicating that playing with mathematical objects (manipulatives, drawings, symbols, pictures and diagrams) is important for children's mathematical development (e.g., Anthony & Walshaw, 2007; Dooley et al., 2014), it is unlikely that many children will develop sophisticated mathematical ideas without some explicit teaching or guided interaction (e.g., Clements et al., 2013). Carbonneau et al.'s (2013) meta-analysis examining the effects of concrete manipulatives on mathematics achievement across all years of schooling and beyond found that high levels of instructional guidance, or explicit teaching, were generally associated with higher effects on outcomes. Hence, Carbonneau et al. argued that, in general, explicit teaching helps learners establish connections between the concrete manipulatives and the intended mathematical ideas, which in turn facilitates understanding. However, this finding is largely based on evidence from studies with older children and we were not able to find evidence within our dataset to support this finding, because we did not have sufficient data on the levels of instructional guidance in the original studies. However, this finding is supported and emphasised by several of the Best Evidence Syntheses (Cross et al., 2009; Clements et al., 2013; Deans for Impact, 2019; Nunes et al., 2009) and, in our judgment, there is relatively good evidence to support the importance of structured teaching to make best use of manipulatives and representations.

There has been a great deal of interest amongst educators in England in Concrete-Pictorial-Abstract (CPA) approaches to the teaching of mathematics, an approach which is broadly supported by research into children's development, although the evidence from the reviews and syntheses of research indicate that this is a more complex process than a simple cycle. The Deans for Impact (2019) report, for example, states that "Young children begin to understand abstract mathematical concepts through concrete representations, and learn to apply what they know in new contexts by gradually transitioning from concrete to visual to abstract" (p.13). However, we found no specific evidence relating to interventions explicitly labelled as CPA, which may be because the notion of CPA broadly is central to interventions involving manipulatives and representations. The Deans for Impact (2019) report highlights the importance of "concreteness fading" to describe how symbols and other abstract representations need to gradually replace concrete representations in their thinking about quantity (see also Fyfe et al., 2014). Teachers and other adults have an important role to play in helping children to make connections between different forms of representation (e.g., Cross et al., 2009). Indeed, Cross et al. (2009) argue that teaching should be directed at helping children to move from concrete to abstract thinking. Key to this, they argue, is the use of a wide range of examples and non-examples, as well as enabling children to link their informal knowledge to formal language, symbols and procedures.

Several interventions involved counting, comparison, estimation, addition and subtraction using objects, number lines and dot patterns (either physical or virtual). One study (Casey et al., 2008) used building blocks to improve children's spatial awareness (as measured by a mental rotation task).

Ramani & Siegler's (2011; 2012) research has demonstrated that playing a linear board game, where children move playing pieces, can be an effective way of developing children's numeracy skills, particularly for children from disadvantaged backgrounds, and replications demonstrate that this also has benefits for other children struggling with number (Deans for Impact, 2019). Although this work is in the context of a particular resource, "The Great Race Game", in our judgment, and the judgment of other experts (e.g., Deans for Impact, 2019), this finding extends to well-known and readily available board games, such as Snakes and Ladders, providing opportunities to develop strategies such as 'counting on'. However, this intervention involves a gaming context as well as the use of manipulatives, and the studies did not examine the extent to which the benefits were due to the manipulation of playing pieces, the gaming context or the combination of the two.

Four of the studies investigated the use of movement or gesture alongside manipulatives and representations, and all showed a positive effects. This finding, that children benefit from actually moving and interacting with manipulatives to understand mathematical ideas, is supported by Cross et al. (2009), who argued that, while pictures are a valuable tool for learning, manipulatives are more effective, because children can manipulate them in ways that physically represent or resemble mathematical concepts, processes and operations. In addition, Cross et al.'s (2009) expert-judgment-based review argued that well-designed virtual manipulatives may be able to offer *more* manipulative flexibility and *more* opportunities for children to describe their actions than is the case with concrete manipulatives. However, we were not able to identify any studies that provided evidence to support this position or any studies that directly compared virtual and concrete manipulatives (see Computer-assisted instruction, apps and technology tools).

There is a great deal of evidence from studies of children's mathematical learning highlighting the importance of representations in the learning of mathematics (see, e.g., Nunes et al., 2009). Indeed, Nunes et al. (2008) observe that representations (e.g., number symbols and diagrams) are fundamental to mathematics in that they "afford manipulations which might otherwise be impossible" (p. 9). However, learning to interpret, coordinate and use these different mathematical representations is far from straightforward, and children need help from educators to do so. Frye et al.'s (2013) What Works Clearinghouse Guidance Report finds moderate evidence that educators need to help "children to link their informal knowledge with formal representations of math[ematics]" (p.111).

Several best evidence syntheses highlight the importance of showing children different representations of number and helping them to make connections between them (Cross et al., 2009; Deans for Impact, 2019; Clements et al., 2013. Cross et al. emphasise that educators should not discourage children's use of fingers, but rather should help children to use fingers *as a representation (or manipulative)* and in ways that help them to develop their understanding of counting, addition and subtraction.

Educators need to consider carefully what manipulatives and representations to use and for how long in order to enable children to develop increasingly sophisticated mathematical ideas (e.g., Cross et al., 2009). One study extracted from Carbonneau et al.'s (2013) meta-analysis strikes a cautionary note regarding manipulatives: Battle (2007) found that Grade 1 children (Year 2) taught addition and subtraction with counters performed very much worse than children taught without counters. A possible cause is that, in this case, the counters encouraged some children to use less sophisticated but familiar strategies (such as 'count-all' rather than 'count-on' methods). This illustrates an argument made by Deeley et al. (2014) that, when misused, manipulatives and representations can inhibit children's development. One way to avoid this is to use a development progression, or learning trajectories, approach to guide teaching (e.g., Clements et al., 2013). Frye et al.'s (2013) What Works Clearinghouse Guidance Report finds moderate evidence to support such an approach, largely from research programmes such as *Building Blocks* (Clements & Sarama, 2007). Carbonneau et al.'s (2013) meta-analysis examined the effect of time, and found some evidence to show that, in general, interventions using manipulatives for up to 45 days had a greater effect than interventions over longer periods. This may be because in longer interventions children may come to rely on concrete manipulatives rather than develop more sophisticated approaches and ideas. However, as Carbonneau et al. caution, there is a need for further research to understand the effects of time (as well as the related issues of frequency, duration and intensity).

Links to other strands:

Play: Play is an important element of young children's learning and it is important that children have opportunities to engage in both free and structured play, with a range of manipulatives. Key to this is creating a rich mathematical environment including a diverse range of manipulatives that provide opportunities for educators to observe, assess and intervene where appropriate.

Explicit teaching: As noted in the findings, in general, children need help from educators in order to develop sophisticated mathematical ideas from manipulatives.

Computer-assisted instruction, apps and technology tools: Many of the CAI and Apps interventions involved the use of virtual manipulatives, although we identified no evidence to indicate the relative efficacy of virtual and physical manipulatives.

Movement and gesture: The evidence suggests that manipulatives and representations are more effective when children actually carry out the *manipulation* themselves in order to explore mathematical ideas.

Evidence base:

We judge the experimental evidence supporting the use of manipulatives and representations in general to be **moderate**.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	2	19 effects from 17 studies, none at scale
Methodological quality of the original studies	2.5	All small scale, some good methodological quality
Consistency of results	1	Heterogeneity
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	2	Supports findings from the original studies
Overall Quality of Evidence judgment	**2**	**Moderate**

Relevance:

The relevance of the evidence is judged to be **moderate.**

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	2	The identified studies were conducted in the US, Belgium, Germany, Netherlands, Sweden and the Netherlands. None in England.
How the interventions were defined and operationalised	2	Varied interventions. However, manipulatives and representations sufficiently 'broad' as an intervention to allow some variation.
Any focus on particular topic areas	2	Mostly, but not exclusively, number, but includes pattern as well as geometry / spatial reasoning.
Age of children /phase of education	2	The studies were carried out across the age range and in contexts that have relevance to both Early Years (8) and Key Stage 1 (7).
Ease of implementation	3	Judged to be relatively straightforward to implement
Overall relevance judgment	**2**	**Moderate**

Whole-curriculum interventions

There is evidence to support the use of eight different whole-curriculum interventions to increase attainment in mathematics. Three of these interventions have been evaluated more than once at scale and have shown positive effects on children's attainment: Building Blocks, Project M^2 and Oxford Mathematical Reasoning. Only one of these, Oxford Mathematical Reasoning, has been designed and evaluated in England. Implementing whole-curriculum interventions at scale can be a challenge.

Definitions:

Whole-curriculum interventions include a range of interventions for whole classes of children covering either the whole or a substantial part of the mathematics curriculum over a period of at least a term. Some interventions, such as *Building Blocks*, focus on the entire mathematics curriculum and cover all of children's mathematical experiences. Other interventions, such as *Oxford Mathematical Reasoning*, focus on providing a series of regular whole-class lessons to be taught over at least a term (10-12 weeks). These are often referred to as 'curriculum interventions' in the US (see, e.g., Pellegrini et al., 2020).

Findings:

We identified 14 studies that evaluated a total of 8 different whole-curriculum interventions that met our inclusion criteria. Four interventions have been evaluated at scale in more than one study: *Building Blocks* (4 studies), *Big Maths for Little Kids* (2 studies), *Oxford Mathematical Reasoning* (2 studies) and *Project M^2* (2 studies). *Building Blocks* and *Oxford Mathematical Reasoning* have each been independently evaluated, whereas *Project M^2* has only been evaluated by teams that included the developers.

Following the approach adopted by Slavin and colleagues' Best Evidence Syntheses (e.g., Pellegrini et al., 2020; Slavin & Lake, 2008), we have aggregated the effects in a meta-analysis. This gave an overall moderate effect (d=0.44, 95% CI: 0.16, 0.72). However, this estimate should be treated with some caution due to differences between the various programmes and, indeed, there was a very high degree of heterogeneity across the effects (I^2=97.9%). There was also some variation in the effects reported for each intervention, which may indicate some variation in implementation or research design; the effects ranged from *d*=0.11 to 1.09 for *Building Blocks*, from *d*=0.30 to 0.99 for *Big Maths for Little Kids,* from *d*=0.08 to 0.20 for *Oxford Mathematical Reasoning,* and from *d*=0.25 to 1.88 for *Project M^2.*

Only one of these whole curriculum interventions, *Oxford Mathematical Reasoning*, has been designed and evaluated in England; the other seven interventions have been developed for use in the US.

Of the eight whole-curriculum interventions, all but two, *Building Blocks* and *Project M^2*, are targeted at Kindergarten or Pre-Kindergarten. *Project M^2* is focused on Kindergarten and Grade 1, whilst *Oxford Mathematical Reasoning* is targeted at Year 2 children.

Building Blocks is a particularly important programme that has been carefully developed and evaluated over the past 20 years in a number of studies led by Doug Clements and Julie Sarama (e.g., Clements & Sarama, 2008). *Building Blocks* is a comprehensive research-based curriculum designed for children in Pre-Kindergarten settings that integrates mathematics throughout the school day and addresses the entire mathematics curriculum (Clements et al., 2011). The intervention provides software, manipulatives, games and texts, together with guidance on whole-class, small-group and individual activities. The intervention is informed by a learning trajectories approach based on an extensive programme of research on children's learning (e.g., Clements & Sarama, 2007). Educators receive substantial professional development (13 days over two years) together with coaching visits. An alternative focused on the *Building Blocks* software intervention has also been evaluated (see Computer-assisted instruction, apps and technology tools strand). One potential criticism of evaluations of Pre-Kindergarten mathematics interventions is that the treatment or intervention group is compared to a business-as-usual control group in which children engage in only very limited mathematics learning. However, in an earlier study, Clements et al. (2011) compared *Building Blocks* to an active control in which new mathematics curricula were introduced and found a large positive effect (g=0.72).

In the What Works Clearinghouse guidance on teaching mathematics to young children, Frye et al. (2013) cite *Building Blocks* extensively to support recommendations that include the use of a research-based developmental progression to teach number and operations, the need for a broad mathematics curriculum (that includes geometry, patterns, measurement and data analysis), the use of progress monitoring, encouraging children to see and describe their world mathematically as well as ensuring both dedicated time for mathematics and that mathematics should be integrated throughout the school day. However, they judge only the first of these recommendations to have anything more than minimal evidential support. Other expert-judgment-based reviews cite *Building Blocks* as evidence for the importance of using developmental progressions or learning trajectories to guide teaching and for integrating play into mathematics teaching (Anthony & Walshaw, 2007; Clements et al., 2013; Dooley et al., 2014). In our judgment, *Building Blocks* is an impressive evidence-based programme which does provide some general support for these findings. However, since it is a comprehensive programme, it is difficult to ascribe causal mechanisms to any one element of the programme. Hence, while these findings and recommendations do appear reasonable on the basis of both correlational studies and also expert judgment, they should nevertheless be treated with some caution.

The evidence from this review suggests that some whole-curriculum interventions may be an effective way to raise attainment. However, some caution needs to be exercised here. First, we note again that only one of the interventions, *Oxford Mathematical Reasoning*, has been designed and evaluated in England and, in the most recent effectiveness trial, the effect size achieved was relatively small (d=0.08). Interventions designed for the US, such as *Building Blocks*, would require considerable adaptation for widespread use in the English context, to align with the mathematics curriculum in England and the pedagogic traditions of English Early Years settings. Second, scaling up such interventions is far from straightforward, and many interventions struggle to achieve high levels of fidelity. The *Building Blocks* programme generally achieves relatively high levels of fidelity (e.g., Clements et al., 2011). This may be because the programme adopts an approach, derived from a review of implementation research, based on principles that include involving stakeholders, being clear about permitted adaptations and creating incentives (Sarama et al., 2008). This approach has shown benefits over implementation without this support (Sarama et al., 2008; see also Clements & Sarama, 2008). However, the high level of fidelity may also in part be due to external support at a school district level. Wang et al. (2016) compared the effects of comprehensive programmes with supplemental programmes, which they argued were likely to be easier to implement, and found no difference in the effects. However, we note that it is possible that the effects of some of the supplemental programmes considered may have been high because they were implemented by the researchers themselves and, hence, achieved good fidelity.

Links to other strands:

Many of the whole-curriculum interventions involve a variety of strategies including explicit teaching, play, etc.

Evidence base:

We judge the experimental evidence supporting the use of whole-curriculum interventions to be **moderate**.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	2.5	14 effects; 3 studies at scale
Methodological quality of the original studies	2	Several well-constructed studies; some variation in the effects reported for each programme evaluated more than once
Consistency of results	1	Interventions are of very different types & approaches
Reporting bias	1	Many interventions evaluated by developers
Evidence from systematics reviews and best evidence syntheses	2.5	Considerable expert / academic support for research-based whole curriculum interventions. One intervention (Building Blocks) used as exemplar in several BES.
Overall Quality of Evidence judgment	**2**	**Moderate**

Relevance:

The relevance of the evidence is judged to be **moderate.**

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	2	Aside from 2 in England, the identified studies were conducted in the US.
How the interventions were defined and operationalised	2	Mostly well-defined & manualised interventions. Some, like Building Blocks, involve holistic intervention.
Any focus on particular topic areas	3	Studies involved both number and/or calculation and geometry / spatial reasoning.
Age of children /phase of education	2	Mostly, but not exclusively, Early Years. England-based studies KS1 (1).
Ease of implementation	2	Some studies, eg Building Blocks, may need external support and pressure to implement.
Overall relevance judgment	**2**	**Moderate**

Feedback and formative assessment

A small number of studies and several best evidence syntheses and expert reviews provide evidence to support the use of feedback and formative assessment practices in teaching mathematics to young children in the Early Years and Key Stage 1. Evidence from one study supports the use of an online tool to enable educators to use their observations of children's mathematical activity to plan future instruction. There is a small amount of evidence to support the use of corrective feedback alongside the opportunity to engage conceptually. Evidence also supports the use of teaching programmes that include formative assessment practices. However, evidence is limited and more research focusing specifically on the use of formative assessment practices with young children is needed.

Definitions:

Feedback is the *provision of information, in any form, to children, or educators, about children's learning and progress. Corrective feedback consists of evaluating, and, if necessary, correcting children's responses.*

Formative assessment refers to processes of gaining insight into learning and the use of those insights to improve learning. The term formative assessment is often used rather loosely to refer to a range of diverse pedagogic strategies (Bennett, 2011). Broadly, these practices are intended to give educators insight into children's understandings, thus enabling educators to plan appropriate mathematical learning opportunities and provide feedback to children.

Findings:

We identified 6 studies that included focus on feedback and formative assessment and which met our inclusion criteria with sufficient data to aggregate effects in a meta-analysis. This gave an overall moderate effect (d=0.31, 95% CI: 0.18, 0.44). Effects in all studies were positive. The studies covered settings relevant to both Early Years and Key Stage 1. However the effects reported must be interpreted with some caution, since, for four of the six studies, feedback and formative assessment were components of the interventions, and one factor among many; the effect size reported is thus for the intervention overall, and not feedback and formative assessment specifically. We therefore report a limited evidence base for the impact of the use of feedback and formative assessment on young children's mathematical learning, consistent with the findings of Cross et al. (2009).

Despite this limited evidence base, best evidence syntheses and expert reviews support the importance specifically of feedback (Cross et al., 2009), and formative assessment practices more generally (Anthony & Walshaw, 2007; Clements et al., 2013; Cross et al., 2009; Dooley et al., 2014), arguing that this is a key element of effective teaching and learning. In particular, these reviews emphasise the importance of educators observing children engaging in mathematics in order to assess how to intervene so as to support them to develop more sophisticated mathematical concepts and skills. However, both Clements et al. (2009) and Cross et al. (2009) argue that educators devote very little time to observing young children engaging in mathematics activities. Clements et al. (2013) suggest that interviews, documentation of children's talk and collating samples of work can be productive ways for educators to understand young children's thinking. Cross et al. (2009) add that these activities, particularly observation, tasks and interviews (involving probing questioning), can be 'rigorous, focused and deliberate' (p.256) and that they should be employed frequently. In the light of this, it is notable that only one study reported below (Polly et al., 2017) expressly investigates the use observation as an assessment method.

Cross et al. (2009, p.255) also note that an understanding of developmental progression in mathematics is needed to enable educators to use assessment to inform teaching, and one study in our set responds to this recommendation. Polly et al. (2017) report the use of an internet-based tool that supports educators in making diagnostic assessments in one-to-one sessions with children. In this study, information arising from educators' observations of children's responses to tasks was entered into a formative assessment tool. This tool included a built-in rubric and provided an assessment based on the responses entered; this then linked to further instructional materials. The authors report greater gains from the use of these formative assessment practices for pupils in disadvantaged school communities compared to business as usual teaching and, consistent with the recommendations of Cross et al. (2009), where assessments were made more frequently.

Corrective feedback is a core component of many direct instruction interventions (see the 'Explicit teaching' strand). Some evidence supports the use of corrective feedback on accuracy in number fact recall and problem solving. However, the limited evidence available suggests that this should be combined with opportunity to engage conceptually. This is illustrated in the difference in design and outcome between two studies, Fuchs et al. (2006) and Popa & Păuc (2015). In the first, Fuchs et al. (2006) addressed recall of addition and subtraction facts in Grade 1 low attainers using computer-assisted flashcard software. Immediate corrective feedback indicated where errors in recall had been made and lengthened the stimulus display time for subsequent examples. Here, a small effect for recall of addition facts only was reported, but with no transfer of learning gains to story problems. In contrast, Popa & Păuc (2015) enacted an intervention specifically focusing on formative assessment practices (which they termed 'dynamic assessment'). In this case, the use of corrective feedback was combined with engaging pupils in reflection on their problem-solving approaches, identifying solution paths, and the use of question prompts to support strategy selection and to encourage autonomy in the analysis of problem-solving processes. Used with high achieving 6- and 7-year olds in Romania, these approaches led to a moderate positive effect.

Two further studies support the use of corrective feedback alongside the opportunity to engage conceptually. Bryant et al. (2011) and Fuchs et al. (2013) both employed a tutoring model for low-attaining pupils in which manipulatives and problem solving were used to develop arithmetic skills (Fuchs et al., 2013) and conceptual understanding of number and calculation (Bryant et al., 2011). In each of these studies, corrective feedback was a part of the intervention, combined with the use of alternative calculation strategies.

Lastly, Sarama et al.'s (2008) study presents an example of an intervention in which assessment practices were part of a much more wide-ranging intervention. This teacher professional development programme and curriculum intervention for Early Years classrooms (US Pre-K) employed both software and class-based activities following a learning trajectory programme. The programme evaluated a wide range of instructional characteristics, including approaches to formative assessment such as the use of scaffolding, listening to children's responses and the use of assessment to adapt children's tasks. Observation of class teachers revealed strong fidelity of implementation, with listening to children and use of scaffolding evaluated as the most consistently implemented of the individually evaluated components of assessment. Thus, whilst the effect size cannot be attributed to formative assessment specifically, its integration into the intervention is in line with the recommendations of best evidence syntheses.

Links to other strands:

Computer-assisted instruction, apps and technology tools: The use of software to support elements of skill rehearsal, to provide corrective feedback and to support educator assessment is noted above.

Executive functions and metacognition: Corrective feedback is a component of many interventions aimed at improving executive functions or metacognition.

Explicit teaching: As noted above, feedback and/or formative assessment practices are included as components of a more comprehensive teaching intervention. In particular, corrective feedback is often a feature of direct instruction interventions.

Individual and small-group tutoring by adults: Formative assessment and feedback are frequently part of a tutoring programme.

Evidence base:

We judge the experimental evidence supporting the use of feedback and formative assessment to be **weak-to-moderate**. This may be surprising given that feedback, in particular, is generally regarded as an intervention with relatively strong evidence. However, much of the evidence relates to older pupils (beyond Key Stage 1).

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	1.5	6 studies. 2 at scale (although not clear whether clustered analysis carried out). Also, for all but 2 of the 6, feedback/formative assessment is part of a holistic intervention.
Methodological quality of the original studies	1.5	Mixed, 2 at scale
Consistency of results	1	Heterogeneity
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	**2**	Considerable expert-based support including reviews and indications from meta-analyses of ways in which feedback may be effective/ineffective
Overall Quality of Evidence judgment	**1.5**	**Weak-to-moderate**

Relevance:

The relevance of the evidence is judged to be **moderate.**

Threat to relevance	Grade	Notes
Where and when the studies were carried out	2	All but one of the studies carried out in the US
How the interventions were defined and operationalised	1	The interventions are not described in detail. For 4 studies, the assessment is part of a wider more holistic intervention and the remaining two focus on formative assessment and dynamic assessment, respectively.
Any focus on particular topic areas	1	Mostly, but not exclusively, number
Age of children /phase of education	2	The studies were carried out across the age range and in contexts that have relevance to both Early Years and Key Stage 1.
Ease of implementation	2	Generally considered to be relatively straightforward to implement, although may require sophisticated approaches to pedagogy.
Overall relevance judgment	**2**	**Moderate**

Mathematical talk and the use of storybooks

There is a consensus that high quality talk and extended discussion is key to children's development, although there is very limited experimental evidence about specific interventions to raise the quality of number-related talk. Relatedly, there is a small but growing body of evidence to support a specific intervention promoting mathematical talk, the use of storybooks. Some research indicates that e-books and educator-led use of storybooks can be effective. Storybooks can help develop children's spatial as well as numerical understanding. Storybooks provide an opportunity to support mathematical talk and discussion. As with any resource, educators need to consider carefully how, and which, storybooks should be used to help children develop more sophisticated mathematical ideas. The intervention evidence is all from Kindergarten, or Early Years, settings.

Definitions:

The evidence on *storybook*s relates to picture books (with or without text) in structured group reading sessions, led by an educator, and to *e-books,* which provide an interactive experience via a computer or tablet in which a narrator reads to a child.

Findings:

Talk and vocabulary: We identified 2 studies which met our inclusion criteria, which is too few to conduct a meta-analysis. Both studies were carried out with Key Stage 1 children. Powell and Driver (2015) examined the effect of a tutoring intervention in which Grade 1 children with mathematical difficulties were taught the meaning of key vocabulary related to number, comparison and addition, compared to children who received the same intervention, but without a specific focus on vocabulary, finding a moderate effect size (d=0.49). Russo & Hopkins (2018) compared two approaches to discussion in the context of challenging tasks with Grade 1 and Grade 2 children: a discussion-first approach and a task-first approach, but did not compare these to a business-as-usual control. They found a positive effect on fluency for the discussion-first approach, but no difference in the effects on problem solving.

These studies examined interventions that specifically addressed talk as the primary feature, although very many of the interventions in our database involved guidance for teachers and other adults on talk, questioning and discussion (e.g., see in particular the whole-curriculum and tutoring interventions). There is agreement across the expert-judgment-based reviews and best evidence syntheses that high quality talk and extended discussion is key to children's development (Anthony & Walshaw, 2007; Cross et al., 2009; Clements et al., 2013; Frye et al., 2013). Dooley et al. (2014, p.37), for example, make the case for mathematical talk as enabling "sustained shared cognitive engagement … ensuring optimal cognitive challenge for all children … [and] mak[ing] their thinking visible". Clements et al. (2009) recommend encouraging children to use informal language to describe mathematical ideas and that teaching should help children to connect their informal knowledge to more formal vocabulary.

The use of storybooks: There is a small but growing body of evidence to support the use of storybooks to teach mathematics, most of it involving the specific use of talk. We identified 6 studies that examined the effect of using storybooks and met our inclusion criteria. There was sufficient data in the papers to aggregate all of these effects in a meta-analysis. This gave an overall large effect (d=0.96, 95% CI: 0.29, 1.63). All of the effects were positive, but the range of effects was great, varying from small (d=0.04) to very large (d=2.0). This diversity suggests that the way in which storybooks are used is critical to their effectiveness. Additionally, this result should be treated with some caution because the large effects may be inflated by aspects of the study design or implementation. Evidence from one best evidence synthesis (Cross et al., 2009) and one expert review (Dooley et al., 2013) provide some relatively weak support for the use of storybooks.

All of the storybook intervention studies were conducted in Kindergarten settings, although, in our judgment and the judgment of several expert reviews (e.g., Dooley et al., 2013), these findings would be expected to generalise to more formal Key Stage 1 settings.

The studies all involved the use of specially chosen or specially designed storybooks. One study (Casey et al., 2008) focused specifically on geometry, whilst others addressed number or mathematics more generally. Two studies examined the use of *Grandfather's Minibus,* an interactive e-book available in Israel, and both found a very large effect (Segal-Drori et al., 2018; Shamir & Baruch, 2012).

The four (physical) storybook interventions all involved guidance to support mathematical talk and educator questioning (Casey et al., 2008; Hassingder-Das et al., 2015; Purpura et al., 2017; van den Heuvel-Panhuizen et al., 2016). In one case (Purpura et al., 2017), notecards were placed at key points in the books, prompting the educator interventions and thus facilitating mathematical discussion. Dooley et al.'s (2013) expert review discusses ways in which educators can engage in discussion with children. Clements et al. (2013, p.38) observe how a storybook can be used to assess, or explicitly teach, mathematical vocabulary such as 'more' or 'fewer'.

Throughout this review, we have cited evidence from Best Evidence Syntheses and other expert-judgment-based reviews, emphasising that educators need to consider what strategies and resources to use and how they should be used, and this also applies to storybooks (e.g., Anthony & Walshaw, 2007; Clements et al., 2013; Cross et al., 2009; Dooley et al., 2014). Based on a review of relevant research and consultation with expert practitioners, van den Heuvel-Panhuizen & Elia (2012) have developed a framework for evaluating the suitability of picture books for children's mathematical development (see also Dooley et al., 2014). More educator-friendly evidence-informed guidance is available from the US-based Development and Research in Early Math Education (DREME) website: https://dreme.stanford.edu/.

Links to other strands:

Many of the interventions addressed in other strands involve guidance on the use of talk by educators. Such guidance is a particular feature of *whole-curriculum* and *tutoring* interventions

Computer-assisted instruction, apps and technology tools: As noted in the findings, there is some evidence to support the use of e-books.

Manipulatives and representations: Mathematical ideas are represented in picture books and, hence, the findings in the manipulatives strand are relevant. When choosing picture books to support mathematical learning, it is important to consider how mathematical ideas are represented in the illustrations and the text.

Evidence base:

We judge the experimental evidence supporting effective interventions promoting high quality talk to be **minimal**. We judge the experimental evidence supporting the use of storybooks to be **weak-to-moderate**.

Talk and vocabulary

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	1	1; 1 additional study did not compare intervention with control
Methodological quality of the original studies	1	All small-scale
Consistency of results	N/A	Too few studies to make judgment; varied interventions
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	1	Supports findings from the original studies about the importance of the development of talk, examples given support (and extend) the findings of the original studies. But little evidence to support actual interventions.
Overall Quality of Evidence judgment	**0**	**Minimal**

The use of storybooks

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	1.5	6, 1 at 'medium' scale [18 schools]
Methodological quality of the original studies	2	Medium scale not a clustered analysis
Consistency of results	1	Heterogeneity
Reporting bias	NK	
Evidence from systematics reviews and best evidence syntheses	1.5	Supports use of picture books, drawing on evidence largely from the experimental studies included in this meta-analysis
Overall Quality of Evidence judgment	**1.5**	**Weak-to-moderate**

Relevance:

The evidence on interventions to promote talk is judged to be of **minimal** relevance. The evidence on interventions to promote the use of storybooks is judged to be of **moderate** relevance.

Talk and vocabulary

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	1	The 2 identified studies were in US and Australia
How the interventions were defined and operationalised	N/A	Too few studies to make judgment
Any focus on particular topic areas	1	Number and geometry / spatial reasoning.
Age of children /phase of education	N/A	Too few studies to make judgment
Ease of implementation	N/A	Too few studies to make judgment
Overall relevance judgment	**0**	**Minimal**

The use of storybooks

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	2	The identified studies were conducted in the US, Israel and Netherlands. None in England.
How the interventions were defined and operationalised	2	Well-defined, although different.
Any focus on particular topic areas	2	Mostly number
Age of children /phase of education	1	All Early Years.
Ease of implementation	2	Some studies included well-defined strategies/questions to use.
Overall relevance judgment	**2**	**Moderate**

Movement and gesture

Although there is a substantial body of literature on gesture and learning mathematics, there is only a small number of recent studies that examine interventions involving movement or gesture. Although these provide some evidence to suggest the value of using movement and gesture to teach mathematics, the evidence base is weak and covers a disparate range of interventions. Nevertheless, movement and gesture is a low-cost strategy that could be used relatively easily. The evidence is relevant to both Early Years and Key Stage 1 settings.

Definitions:

In this strand, we refer to interventions where *movement* or *gesture* by children is a central component of the intervention. We have not included linear board games or strategies involving fingers, which are both coded as manipulatives and representations.

Findings:

We identified 4 studies that examined the effect of using physical movement and/or gesture and met our inclusion criteria. Since the studies examine the effects of different interventions, the effects are not aggregated. In all cases, movement or gesture was combined with other strategies, such as verbalising.

All but one of the studies was small scale; the exception being Have et al.'s (2018) study involving 12 schools. Nevertheless, all the studies showed a similar effect size (*d*=0.37 to 0.40), except for one (*d*=0.91, Jordan et al., 2015). The studies examined a range of strategies involving movement and gesture: physically moving along a large floor number line (Ruiter et al., 2015); and gesturing to indicate a set whilst counting or comparing numbers of objects on a computer (Jamalian, 2015). Two studies investigated interventions where teachers integrated mathematically-related physical activity into mathematics lessons over an academic year after receiving professional development (Have et al., 2018; Shoval et al., 2018). A further study, not included in the meta-analysis, involved an element within a wider computational thinking intervention: instructing another child to move along a number line in the context of Scratch programming (Sung et al., 2017).

There is a substantial literature examining how gesture contributes to learning (e.g., Goldin-Meadow, 2011) and evidence from three best evidence syntheses (Cross et al., 2009; Clements et al., 2013; Frye et al., 2013) and two expert reviews (Anthony & Walshaw, 2007; Dooley et al., 2013) provide support for the use of movement and gesture during play and in the context of both numeracy and geometric tasks.

Links to other strands:

Manipulatives and representations: As noted in the manipulatives strand, children benefit from actually moving and interacting with both physical and virtual manipulatives.

Evidence base:

We judge the experimental evidence supporting the use of movement and gesture to be **weak**.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	1	4, 1 at 'medium' scale [12 schools]
Methodological quality of the original studies	2	Some good. Medium-scale study uses clustered analysis
Consistency of results	1	Quite disparate range of studies
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	1	Expert judgment and best evidence syntheses highlight the importance of movement and gesture
Overall Quality of Evidence judgment	**1**	**Weak**

Relevance:

The relevance of the evidence is judged to be **weak.**

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	1	Denmark, Israel & US.
How the interventions were defined and operationalised	1	Range of different approaches
Any focus on particular topic areas	1	Mostly number
Age of children /phase of education	1	The studies were carried out across the age range and in contexts that have relevance to both Early Years (2) and Key Stage 1 (2).
Ease of implementation	1	Unclear, would require description and PD.
Overall relevance judgment	**1**	**Weak**

Peer tutoring and cooperative learning

There is some weak evidence from two studies indicating some benefits for peer tutoring with young children, although both studies consider one specific peer-tutoring intervention. There is only a very limited amount of evidence addressing cooperative learning with young children.

Definitions:

Cooperative learning refers to interventions, usually structured, where children work with peers to solve a common mathematical problem or achieve a mathematical goal, usually with the explicit aim of developing wider mathematical, social and/or communication skills and capabilities.

Peer tutoring refers to interventions in which a same-age or older child coaches, or tutors, another child, although in the Peer-Assisted Learning Strategies (PALS) intervention children take turns to tutor each other.

Findings:

Peer tutoring: We identified 2 small-scale studies that examined the effect of peer-tutoring and met our inclusion criteria. The studies, from 2001 and 2002, both examined the effects of one specific peer-tutoring intervention: Peer-Assisted Learning Strategies (PALS). In PALS, same-age peers work in pairs and take turns to coach each other. The interventions were evaluated by the development team in relatively small-scale studies and found small to medium effects in favour of PALS of d=0.32 for Kindergarten and 0.17 for Grade 1.

Cooperative Learning: We identified 2 small-scale studies that examined the effect of cooperative learning and met our inclusion criteria. The effects were judged to be too varied to aggregate. The more recent study found no effect for cooperative learning compared to both individual learning and to a control group (Meloni et al., 2017). An older study found cooperative learning to have a large effect compared to a business as usual control group (d=1.08, Tarim, 2009).

The best evidence syntheses and expert-judgment-based reviews did not directly address either cooperative learning or peer tutoring, although there was implicit support for group-based play and collaborative activities (e.g., Clements et al., 2013; Dooley et al., 2014).

Links to other strands:

No specific links.

Evidence base:

We judge the experimental evidence supporting the use of peer tutoring to be **weak** and cooperative learning to be **minimal**.

Peer tutoring

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	1	2; plus 2 further with insufficient data to be included
Methodological quality of the original studies	1	Both small scale studies and both (PALS intervention) evaluated by the developers
Consistency of results	N/A	Too few studies to make judgment
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	0	No evidence identified
Overall Quality of Evidence judgment	**1**	**Weak**

Cooperative learning

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	0	0 studies, 4 excluded due to insufficient information
Methodological quality of the original studies	N/A	N/A
Consistency of results	N/A	N/A
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	1	Some expert-based support for collaborative activities.
Overall Quality of Evidence judgment	**0**	**Minimal / insufficient evidence**

Relevance:

The evidence on peer tutoring is judged to be of **weak** relevance to Early Years and Key Stage 1 mathematics teaching in England. The evidence on cooperative learning is judged to be of **minimal** relevance.

Peer tutoring

Threat to relevance	Grade	Notes
Where and when the studies were carried out	1	The PALS studies both carried out in the US
How the interventions were defined and operationalised	N/A	Too few studies to make judgment
Any focus on particular topic areas	N/A	Too few studies to make judgment; varied interventions
Age of children /phase of education	N/A	Too few studies to make judgment
Ease of implementation	N/A	Too few studies to make judgment
Overall relevance judgment	**1**	**Weak**

Cooperative learning

Threat to relevance	Grade	Notes
Where and when the studies were carried out	1	Studies carried out in Italy and Turkey
How the interventions were defined and operationalised	N/A	Too few studies to make judgment
Any focus on particular topic areas	N/A	Too few studies to make judgment; varied interventions
Age of children /phase of education	N/A	Too few studies to make judgment
Ease of implementation	N/A	Too few studies to make judgment
Overall relevance judgment	**0**	**Minimal**

Play

A substantial amount of young children's time is spent engaging in play and play-based activities. Broad research emphasises the integral nature of play in child development, but there is a lack of robust evidence of the precise impact of play interventions on learning mathematics. Young children should be encouraged to engage in both free and structured play. By planning and creating rich learning environments, teachers help stimulate children's exploration, discussion and problem solving. Educators can also balance opportunities for free and more structured play with explicit teaching or tutoring with specific mathematical learning goals. Through observing children's play, educators will identify "teachable moments" (Cross et al., 2009) in which they can intervene to discuss salient points, encourage problem solving or connect the play scenario to curriculum topics. Educators can engage in and model mathematical tasks displaying their enjoyment, and this can motivate children to engage in similar behavior (Dooley et al., 2014).

Definitions:

The definition of play can be contentious. In the context of this review, *play* is activity that is not predominantly directed by externally set goals, but instead is defined by the processes that children engage in rather than a specific outcome. Play can be self-chosen and self-directed (*free play*), but can also include structure and rules which may be set by an educator (*structured or guided play*). Play can include many types of content and behaviours, including:

Pretend play: using imagination to role play, such as going to the shop,
Object play: manipulating and building with blocks or jigsaws*, and*
Social play: playing in small groups to solve a problem or puzzle.

Findings:

Through the review process, we identified two studies that investigated the impact of solely play-based interventions on learning and met our inclusion criteria, only one of which reported an effect; therefore, meta-analysis was not conducted for this strand. The one reported effect size was small (d=0.19). Three further studies, reported in the "Whole-curriculum intervention" strand, indicate positive effects for *Building Blocks*, an intervention involving both explicit teaching and play (Sarama & Clements, 2009), although these studies do not investigate the specific impact of the play element. Therefore, in contrast to other strands, there is relatively little robust and direct evidence from interventions on the impact of play, or strategies to encourage mathematically-focused play, on learning mathematics. However, there is a large body of research on the importance of play to children's development and learning and, in our judgment, play is important to young children's mathematical learning. Hence, due to the lack of experimental evidence, we used *Best Evidence Syntheses* to provide further information on the importance of play for learning.

An important aspect of play is that it harnesses children's self-motivation. Colliver (2018) observed that if parents or educators modelled mathematics problem-solving activities near to where children were engaging in other types of play, children were more likely to incorporate these types of activities in their own free play when compared to a control group.

Importantly, these activities were everyday problems that children might encounter, such as measuring out liquids or working out how many blocks they may need to build a wall. Children in the "modeling" intervention group outperformed children in a control group in their numeracy skills at post-test. This study emphasises the important role that the adult fulfills in providing activity ideas and structure, but also the essential role of self-motivation of children to engage in this type of play. By spending time doing these activities adults implicitly communicated the value of mathematical problem solving to children.

Vogt et al. (2018) assessed the effectiveness of an intensive educator-led training intervention using card games and board games (24 × 30 minute sessions). These games focused on skills such as quantity comparison, counting and digit recognition. A second group experienced a play-based intervention, in which children were given the same materials and the same number of sessions as the educator-led group, but instead children were allowed to choose their playing partner and to have free choice of what games to play. Educators in the play-based intervention were asked to introduce the games and support the children in their game-playing, following guided play principles (Weisberg, Hirsh-Pasek, Golinkoff, Kittredge & Klahr, 2016). Results indicated that children in both the educator-led and play-based groups had similar levels of mathematical achievement after the 8-week intervention. These two groups performed better than a business-as-usual control group, indicating that guided, or structured, play scenarios can be effective for children's learning, but children need to be provided with sufficient materials, structure and educator guidance to enable learning.

Best evidence reviews emphasise the importance of play in mathematical learning, especially in relation to providing children with opportunities to think about mathematical concepts and use mathematical language (e.g., Clements et al., 2013). Play is regularly mentioned in relation to geometry and spatial thinking, with building, or construction, blocks play being highlighted as particularly useful for developing spatial awareness and knowledge of shapes (Cross et al., 2009).

Social play provides opportunities for children to engage in creative, imaginative experiences that help to develop functional mathematics skills, such as counting items of food or paying at a shop. Importantly, Cross et al. (2009) highlight that this complex play provides children with the opportunity to develop their self-regulation and executive functions, by controlling their own behavior in complex social situations. Many correlational studies emphasise the importance of these cognitive skills for mathematical learning (e.g., Blair & Razza, 2007). These types of play also provide educators with opportunities to encourage the development of complex language use.

Play provides opportunities for practice of skills that require modelling or scaffolding by adults. By using play-based activities, educators can tap into children's intrinsic motivation to increase attention and remain engaged in challenging tasks. Scaffolded play using board games has been suggested to be particularly beneficial. Board games provide children with a focused play activity and the opportunity for educators to link multiple representations, use complex mathematical language and encourage flexible strategy use (Clements et al., 2003, Deans for Impact, 2019). Educators should feel confident in intervening in these play activities to ensure that opportunities are taken to "mathematize" the experience (Cross et al., 2009), perhaps by pointing out mathematical aspects of objects that children are playing with or encouraging children to solve a social problem using mathematical concepts.

Links to other strands:

Computer-assisted instruction, apps and technology tools: Many Apps involve some element of play, although we did not find studies that investigated this.

Manipulatives and representations: It is difficult to separate out this strand from "Manipulatives and representations", particularly in relation to the use of "snakes and ladders" type linear board games. These games provide structured activities for play, but also evidence suggests that they strengthen and build children's representations.

Whole-curriculum interventions: Play is a key component of *Building Blocks*, an intervention where explicit teaching is based on "finding the mathematics in, and developing mathematics from, children's activity" (Clements & Sarama, 2007, p.138).

Evidence base:

We judge the experimental evidence supporting the use of play-based interventions to be **weak**, although in our judgment, play can make an important contribution to young children's mathematical development.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	0	1 study with sufficient data to meta-analyse original studies, 2 additional studies; none at scale
Methodological quality of the original studies	1	Two small scale, one medium (N=329)
Consistency of results	N/A	Too few studies to make judgment
Reporting bias	N/A	
Evidence from systematics reviews and best evidence syntheses	2	"Theoretical" and expert-based support for play.
Overall Quality of Evidence judgment	**1**	**Weak**

Relevance:

The relevance of the evidence is judged to be **minimal.**

Threat to relevance	Grade	Notes
Where and when the studies were carried out	N/A	Too few studies to make judgment
How the interventions were defined and operationalised	N/A	Too few studies to make judgment
Any focus on particular topic areas	N/A	Too few studies to make judgment
Age of children /phase of education	1	All Early Years (not Key Stage 1)
Ease of implementation	N/A	Too few studies to make judgment
Overall relevance judgment	**0**	**Minimal**

Executive functions and metacognition

A small number of studies have been identified that focus on executive functions and metacognition. The research suggests that educators should encourage children to explain their problem-solving strategies either to themselves or to others. This enables them to gain insight into their own thinking and provides opportunities for children to learn from one another, guided by a teacher or other adult. Educators can also highlight and show children efficient ways to remember and process information and then support them to put these strategies into action. Although it is clear that working memory training does not lead to gains in mathematical achievement, there is some evidence that embedding mathematical content in memory games may be a useful way to encourage children's learning. Educators should remember that children have a suite of sophisticated executive function skills that they can use in the classroom. Games and tasks can be designed to harness and challenge these skills, to boost their development and create opportunities for learning.

Definitions:

To be able to successfully complete a mathematical task, children must be able to regulate their own behaviour, focus their attention, plan, store and process information. The specific skills may be described as metacognition, executive functions or self-regulation.

Executive functions are complex cognitive skills that are associated with learning and behaviour. Specific executive functions include *working memory* (i.e., storing and manipulating information in your mind), *inhibition* (i.e., holding back responses and ignoring distracting information) and *attentional control* (i.e., focusing and shifting your attention as and when required). Executive functions are related to learning through many different pathways: regulating social behaviour enables children to work in groups, monitoring problem solving (metacognition) enables children to learn from their errors and understand mathematics more fully; choosing and switching between different strategies enables children to complete multi-step problems. This may also be referred to as *self-regulation.*

Metacognition is the ability to reflect on and control one's own thinking processes. This reflection enables children to think about the processes or strategies they may use to solve a problem. Executive functions may be useful in multiple domains, such as forming social relationships, concentrating in class or completing complex problem-solving.

The relationship between executive functions, metacognition and self-regulation is complex and contested (Gascoine, Higgins & Wall, 2017). For the purposes of this review, executive functions refer specifically to working memory, inhibition and attentional control, whereas metacognition refers to more general strategies, such as monitoring their mathematical activity and reflecting on or explaining their approaches and strategies. See Muijs and Bokhove's (2017) evidence review for further discussion of metacognition and self-regulation.

Findings:

The review identified four papers that met our inclusion criteria and focused on classroom interventions in mathematics education involving executive functions or metacognition.

Effect sizes could only be extracted from three of the studies, with effects ranging from *d*=0.20 to 1.33. The studies were judged not to be sufficiently conceptually similar to aggregate a meaningful effect size. Therefore, there is little robust evidence on the impact of focusing on metacognition and/or self-regulation on learning. *Best Evidence Syntheses* also provide little evidence for this strand.

Many studies provide correlational evidence that executive functions and metacognition are associated with learning outcomes. That is, more advanced cognitive skills are related to better mathematical achievement. However, there are very few intervention studies that test whether the relationship between these skills and achievement is causal and, more importantly, whether teaching interventions can modify or improve executive functions in ways that in turn improve mathematics learning and/or achievement.

Metacognition

We identified one study that solely focused on metacognition. Rittle-Johnson, Saylor and Swygart (2008) assessed the impact of children explaining correct solutions to pattern-making tasks, either to themselves or to their mothers, compared to a control group who simply repeated the task. There was no difference in outcome between children who explained to themselves or explained to their mothers, but gains in problem solving were observed in both groups when compared to the control group. This study suggested that the process of reflecting on the solution process and explaining how the answer was achieved increased children’s problem-solving skill and, hence, provides some evidence that this metacognitive activity, if prompted, could be an effective way to teach more complex problem solving. (See also, Baten, Praet and Desoete, 2017, discussed below, for a study combining metacognition with numeracy training.)

Combined executive function and numeracy training

A large body of research has established the close association of executive functions, in particular working memory, with success in mathematics (e.g. Deans for Impact, 2019). This has led to significant interest in the potential benefits of executive function training for mathematical achievement, with the majority of studies specifically focusing on working memory training. Evidence suggests that working memory training can improve performance on working memory tasks. However, these studies indicate that working memory training does not improve academic achievement (Melby-Lervag, Redick & Hulme, 2016).

A more nuanced approach to working memory training has recently emerged. Even though executive functions, such as working memory, are important for mathematical achievement, subject-specific knowledge is also essential (Clements et al., 2013). Therefore, it has been suggested that, by combining working memory and numeracy training, real benefits for mathematical achievement may be achieved.

We identified two studies that combined working memory training with numeracy content. Kroesbergen, van’t Noordende and Kolkman (2012) compared the impact on basic number skills of general working memory training and working memory training that used number-based games. In the general working memory training group, games focused on verbal and spatial working memory, such as remembering word lists or the spatial location of objects. The number-based working memory training used similar games, but these had embedded numerical content, such as remembering lists and numbers of items, matching quantities on hidden cards and playing a number line board game. Both groups improved their working memory and number skills after the intervention, when compared to a business as usual control group. Therefore, no unique benefit of embedding numeracy content in working memory training was observed.

Nemmi et al. (2016) compared the impact of number line training (i.e., estimating positions of numbers on a blank number line), working memory training (i.e., remembering the location of items in a visual scene) or combined working memory and number line training on mathematical achievement. Only children in the combined training group displayed significant gains in performance over the business as usual control group after the intervention period. This study may indicate the importance of combining these two types of training to boost mathematical achievement. The authors also emphasise the large differences between children in their response to the intervention, and how this may need to be taken into account in future studies or in practice.

An additional study by Baten, Praet and Desoete (2017) investigated the impact of training that combined metacognitive techniques and mathematical content. This was delivered through a computerised programme. The programme included a number of games; for example, one game required children to remember sequences of digits and quantities. The training supported children in learning *how* to remember mathematical information efficiently and effectively and also provided feedback, so they could learn from their mistakes. This intervention showed positive gains in calculation skills after the intervention, but we do not know whether this was due to the metacognitive or mathematical elements or the combination of the two.

Links to other strands:

Play: Play can provide opportunities for children to practise using metacognitive skills, such as self-explanation. In addition, play places high demands on self-regulation and executive functions; therefore, by engaging in play-based activities children may improve their use of these skills.

Feedback and formative assessment: Timely and appropriate feedback should provide children with the opportunity to reflect on their strategies for problem solving. By using children’s metacognitive skills, educators should be able to increase children’s insight into appropriate strategy use.

Evidence base:

We judge the experimental evidence for this strand to be **minimal.**

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	0	0; 2 studies with ES, but not sufficiently similar to aggregate; 2 additional with insufficient data to aggregate, 2 additional studies examining integrated numeracy (Number Line) and WMT
Methodological quality of the original studies	1	Small experimental studies, one single case design
Consistency of results	1	Inconsistency of definition
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	**2**	Correlational support for an association, but little evidence to support actual interventions
Overall Quality of Evidence judgment	**0**	**Minimal**

Relevance:

The relevance of the evidence is judged to be **minimal.**

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	2	The identified studies were conducted in the US, Belgium, Sweden, Iran and the Netherlands. None in England.
How the interventions were defined and operationalised	1	Similar definitions across studies are used, but varied interventions.
Any focus on particular topic areas	1	Studies either focused on problem solving or more general mathematical achievement. No geometry / spatial reasoning.
Age of children /phase of education	2	The studies were carried out across the age range and in contexts that have relevance to both Early Years (3) and Key Stage 1 (2).
Ease of implementation	1	Little evidence to support actual well-described interventions.
Overall relevance judgment	**0**	**Minimal**

Parent and family numeracy programmes

There is very little evidence about the effectiveness of parent and family numeracy programmes.

Findings:

There is some correlational evidence to indicate that parental interest and engagement in children's mathematics is associated with increased attainment (e.g., Cross et al., 2009), although a recent review suggests that the results are mixed (Napoli & Purpura, 2018). Moreover, there is very little evidence about effective ways of increasing parental engagement that have an impact on children's attainment. Two relatively recent reviews carried out by Greg Brooks and colleagues (Brooks et al., 2008; Cara & Brooks, 2012) identify very little robust evidence about family numeracy programmes. We identified only two relevant studies in our searches, although both were very small scale. Cheung & McBride (2017) found a positive effect for an intervention in China encouraging parents to play numerical board games with their children, whilst Colliver's (2018) intervention encouraging mathematical play between parents and children showed a *negative* effect on learning. Nevertheless, two expert-judgment-based reviews address parental involvement. Cross et al. (2009) cite evidence indicating that parents spend very little time on mathematics with their children and suggest ways in which families can engage in numeracy and other mathematical activities. Dooley et al. (2014) place a great deal of emphasis on developing a partnership between educators and parents about children's mathematics. However, as noted above, there is very limited evidence about *how* to intervene to support effective parental involvement in children's numeracy.

Evidence base:

We judge the experimental evidence supporting interventions or strategies to support transitions to be **minimal**.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	1	2
Methodological quality of the original studies	1	Very small scale
Consistency of results	N/A	Too few studies to make judgment; varied interventions
Reporting bias	N/A	Not known
Evidence from systematics reviews and best evidence syntheses	1	"Theoretical" and expert-based support for working with parents, although not for specific interventions. Correlational studies show mixed evidence
Overall Quality of Evidence judgment	0	Minimal

Relevance:

The relevance of the evidence is judged to be **minimal.**

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	N/A	Too few studies to make judgment
How the interventions were defined and operationalised	N/A	Too few studies to make judgment
Any focus on particular topic areas	N/A	Too few studies to make judgment
Age of children /phase of education	N/A	Too few studies to make judgment
Ease of implementation	N/A	Too few studies to make judgment
Overall relevance judgment	**0**	**Minimal**

Problem solving

There is very little evidence about the effectiveness of problem solving, although we consider that educators should encourage children to engage in reasoning and to tackle non-routine and challenging tasks.

Findings:

In an earlier review of mathematics teaching at Key Stages 2 and 3, Hodgen et al. (2018) identified a body of work that focused on the use of problem solving. However, in the searches for this current review, we identified few studies that explicitly addressed problem solving. In fact, despite a strong focus on challenge in several whole-curriculum interventions, only *Oxford Mathematics Reasoning* explicitly mentioned problem solving (see Worth et al., 2015). One potential reason may be that the term 'problem' is sometimes used simply to indicate a task, such as (in the US) any kind of 'word problem', whereas in other contexts it implies a task with a certain degree of challenge. Hence, the *Building Blocks* intervention does refer to 'problem solving' but includes it as just one type of activity within the overarching theme of mathematising (Clements et al., 2011). For standard word problems, Hembree's (1992) meta-analysis finds that representations and manipulatives are helpful for young children, although he also finds that the value of these problems per se increases with children's cognitive development. However, there is a great deal of support from the best-evidence syntheses and expert-judgment-based reviews for the value of children setting their own problems, for educators ensuring an appropriate level of mathematical challenge for children and for problems as a vehicle for assessment and learning (see, e.g., Clements et al., 2013; Cross et al., 2009, Deans for Impact, 2019; Dooley et al., 2014; Frye et al., 2013). In our judgment, rather than referring to 'problem solving' per se, it may be better to emphasise the characteristics of productive 'problem solving', such as encouraging children to reason, engaging children in non-routine tasks and setting children suitable challenges.

Evidence base:

We judge the experimental evidence supporting the use of problem-solving to be **minimal**.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	N/A	
Methodological quality of the original studies	N/A	
Consistency of results	N/A	
Reporting bias	N/A	
Evidence from systematics reviews and best evidence syntheses	1	Theoretical and expert-based support for use (and importance) of problem-solving.
Overall Quality of Evidence judgment	**0**	**Minimal**

Relevance:

The relevance of the evidence is judged to be **minimal.**

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	N/A	Too few studies to make judgment
How the interventions were defined and operationalised	N/A	Too few studies to make judgment
Any focus on particular topic areas	N/A	Too few studies to make judgment
Age of children /phase of education	N/A	Too few studies to make judgment
Ease of implementation	N/A	Too few studies to make judgment
Overall relevance judgment	**0**	**Minimal**

Professional development and teacher (educator) knowledge

There is widespread consensus on the need for high-quality professional development. However, there is very limited evidence on effective approaches to professional development. Some evidence from expert-judgment-based reviews suggests that coaching together with instructional feedback may be effective, particularly for manualised, or well-described, interventions. Additionally, professional development should address educators' knowledge in three areas: content knowledge of mathematics, pedagogical content knowledge in mathematics and knowledge of children's mathematical development.

Definitions:

Content knowledge (CK) is used to refer to an educator's knowledge of mathematics, more specifically for this review, the mathematics in the Early Years and Key Stage 1 curricula.

Pedagogical content knowledge (PCK) is a term originally coined by Lee Shulman to refer to the knowledge a teacher (or other educator) needs to teach a particular subject area above and beyond the usual content knowledge that a non-educator in that domain would have. In mathematics, this may include some knowledge about connections between mathematical concepts (e.g. Askew et al., 1997), knowledge of children's development and likely difficulties or misconceptions, different approaches to solving problems or tasks, and knowledge of how classroom tasks connect to mathematical ideas (e.g., Baumert et al., 2010).

Findings:

In an earlier review focused on Key Stages 2 and 3 mathematics, Hodgen et al. (2018) highlighted a paucity of evidence relating to effective professional development for teachers. The evidence for Early Years and Key Stage 1 educators is even weaker. Nevertheless, there is widespread consensus amongst both professional experts (e.g., ACME, 2016) and researchers (e.g., Dooley et al., 2014, on the need for high-quality professional development). Few studies focus exclusively on the effects of professional development without an accompanying teaching intervention to be implemented with young children. As a result, it is difficult to isolate and, thus, assess the specific impact of professional development. There is also very limited direct evidence on the effect of teacher (or other educator) mathematical knowledge (or pedagogical content knowledge) on young children's learning. However, our review does provide some evidence relating to professional development and teacher knowledge.

First, several of the best evidence syntheses and expert reviews all point to the importance of professional development and suggest ways in which professional development is likely to be more effective. However, all draw heavily on studies with teachers of older children to support their arguments, and hence these findings are largely based on expert inferences from research rather than being directly supported by either correlational or experimental evidence. Dooley et al. (2104), for example, highlight lesson study and the importance of developing teachers' mathematical knowledge. Clements et al. (2013) argue that often professional development for the teachers of young children is "too unfocused, too superficial, too brief, too sporadic, and without adequate support or follow through" (p.28). To counter this, professional development for teachers of young children should focus on integrating mathematics knowledge for teaching, the psychology of mathematical development and mathematical pedagogy (or didactics). Clements et al. place considerable emphasis on the role of coaching, real-time corrective feedback and the benefits of using learning trajectories to support teachers' knowledge. They also argue for the importance of learning from interventions that have been successfully scaled up, such as *Building Blocks*, which takes a research-informed approach to professional development and coaching (Sarama et al., 2007). Cross et al. (2009) cite Yoon et al.'s (2007) meta-analysis of the impact of teacher professional development on attainment, which found that the effective professional development programmes that they reviewed averaged 53 contact hours in a period of 4 months to a year, which is substantially more contact time than typically available to Early Years teachers. Only one of the identified studies, Cognitively Guided Instruction (Carpenter et al., 1989), addresses mathematics in settings relevant to Early Years or Key stage 1. However, substantial contact time in and of itself does not appear to be sufficient for effective professional development (and nor does it appear to be strictly necessary, judging by the PD contact time in several of the interventions identified for this review). As we have already noted, a large number of the interventions in this review included professional development and/or coaching for teachers or other adults. In some cases, this professional development was relatively substantial (e.g., for *Building Blocks*, teachers received 13 days of training plus individual coaching), but, in other cases, face-to-face training was relatively limited. One common feature of several effective programmes was coaching with instructional feedback (see also, Kraft et al.'s, 2018, meta-analysis of coaching for educators).

Second, although not included in our main database, three studies that we identified in our searches directly address professional development. Sarama et al. (2007) investigated the effects of a scale-up programme for *Building Blocks*, the TRIAD approach, which involved professional development and coaching coupled with strategies to encourage support from school principals on fidelity to the programme. They found evidence to support the TRIAD approach, although this was not isolated from the effects of the *Building Blocks* intervention. Piasta et al. (2015) compared Early Years educators who were randomly assigned to three conditions, each with 64 contact hours: one of two intervention groups addressing either mathematics or science knowledge, using an approached based on Hirsch & Wiggins's (2009) core knowledge curriculum, or an active equivalent time control group focused on arts and creativity education. Children taught by participants in the science group made gains compared to the control, but children taught by those in the mathematics group did not. This suggests that solely addressing mathematics knowledge may not be sufficient and that, as Clements et al. argue, effective professional development needs to integrate mathematics knowledge with knowledge of children's mathematical development and of effective mathematical pedagogy. It is not sufficient simply for educators to have mathematical knowledge; in order to use this knowledge in the classroom they need knowledge of how children learn and the errors they make, as well as of how to intervene pedagogically to support young children's mathematical development (see also, Cross et al., 2009).

Finally, some research suggests that access to some expertise in mathematics education is an important factor in successful professional change (e.g., Millet et al., 2004; see also Spillane, 1999). Robinson-Smith et al. (2018) evaluated the *Maths Champions* programme, which supports a mathematics expert practitioner in Early Years settings to design and implement an action plan for improving mathematics teaching. This evaluation found a small effect for the programme. However, it is possible that this effect may be related to the focus of the intervention on the mathematics action plan. A wider and more substantial programme to support mathematics expert teachers in all primary schools was introduced in England in 2010 as a result of the Williams (2008) Independent Review of Mathematics Teaching in Early Years Settings and Primary Schools. Intended to eventually address mathematics expertise in all primary schools, the programme involved extended university-led training and, initially, substantial financial incentives to teachers. However, despite widespread support for the programme, an independent evaluation found no effects on children's attainment, although there were effects on children's confidence, but mainly in the participants' own classes (Walker et al., 2013).

In summary, although the evidence base is weak, it is possible to draw some tentative conclusions about professional development for educators of young children (and other Early Years practitioners). Professional development contact time does matter, but it is not sufficient. This review suggests that designing effective professional development is not straightforward and that coaching together with feedback may be important. In our judgment, there is some justification for Clements et al.'s (2013) call for professional development that integrates knowledge of mathematics, of children's mathematical development and of mathematics pedagogy.

Links to other strands:

Professional development is a feature of most interventions in this review.

Evidence base:

We judge the experimental evidence supporting the findings for both professional development and teacher (or other educator) mathematical knowledge to be **minimal**, although, in our judgment, professional development is a key component to enabling educators to use strategies or implement interventions effectively.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	0	No studies
Methodological quality of the original studies	N/A	N/A
Consistency of results	N/A	N/A
Reporting bias	N/A	N/A
Evidence from systematics reviews and best evidence syntheses	1	Consensus on the need for high-quality PD, but less consensus on what constitutes high-quality PD. Some expert support for instructional coaching
Overall Quality of Evidence judgment	**0**	**Minimal**

Relevance:

The relevance of the evidence is judged to be **minimal.**

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	N/A	Too few studies to make judgment
How the interventions were defined and operationalised	N/A	Too few studies to make judgment
Any focus on particular topic areas	N/A	Too few studies to make judgment
Age of children /phase of education	N/A	Too few studies to make judgment
Ease of implementation	N/A	Too few studies to make judgment
Overall relevance judgment	**0**	**Minimal**

Transitions

There is very little evidence about approaches, or interventions, relevant to supporting children's transition in mathematics from Early Years to Key Stage 1, or from Key Stage 1 to Key Stage 2.

Findings:

We identified no studies of interventions relevant to the transition from Early Years to Key Stage 1, or from Key Stage 1 to 2, that met the criteria for inclusion in our reviews. Several of the expert-judgment-based reviews argue that in order to facilitate successful and productive transitions, educators need to communicate about children's mathematical learning and to adopt coherent approaches to teaching across the phases (e.g., Dooley et al., 2014; see also Geudet et al., 2016). Clements et al. (2013) note that the effects of mathematics interventions tend to fade away over time and argue that it is therefore important to explicitly follow through on these programmes in more formal schooling at Grade 1 (Year 2) and beyond. However, as Verschaffel et al. (2017) observe, although there is widespread consensus amongst policymakers, professionals and academics on the importance of greater alignment between phases, there is very little evidence about the benefits of this.

Evidence base:

We judge the experimental evidence supporting interventions or strategies to support transitions to be **minimal**.

Aspects of quality of the body of available evidence	**Grade**	**Notes**
Number of original studies	0	0
Methodological quality of the original studies	N/A	N/A
Consistency of results	N/A	N/A
Reporting bias	N/A	N/A
Evidence from systematics reviews and best evidence syntheses	1	"Theoretical" and expert-based support for 'consistency' across transitions. Correlational studies with older children.
Overall Quality of Evidence judgment	**0**	**Minimal**

Relevance:

The relevance of the evidence is judged to be **minimal.**

Threat to relevance	**Grade**	**Notes**
Where and when the studies were carried out	N/A	Too few studies to make judgment
How the interventions were defined and operationalised	N/A	Too few studies to make judgment
Any focus on particular topic areas	N/A	Too few studies to make judgment
Age of children /phase of education	N/A	Too few studies to make judgment
Ease of implementation	N/A	Too few studies to make judgment
Overall relevance judgment	**0**	**Minimal**

Methodology

This was a rapid evidence review drawing on the literature related to Early Years and Key Stage 1 mathematics in England. In many other educational systems, these are termed Pre-kindergarten / Kindergarten and Grade 1, respectively.

Research Question and Aims

This review was commissioned in order to inform the production of an evidence-based guidance report for educators, schools and other Early Years educational settings. The review aimed to answer the following research question:

What is the evidence on the effectiveness of classroom-based interventions for improving mathematical learning of children in Early Years and Key Stage 1 settings?

For the purposes of this review, interventions are defined as changes to existing classroom practice (Simms et al., 2019), which are sufficiently well-described to be implemented by educators in the Early Years or Key Stage 1. In response to specific queries raised by the panel responsible for writing the guidance, we additionally examined evidence about interventions that went beyond the strict classroom-based focus in three areas: Professional development and teacher (or educator) knowledge; Interventions to support transitions; and Parent and family numeracy programmes. We found no evidence relating to a further intervention: grouping by attainment.

A Rapid Review

While methods for rapid reviews vary significantly, our approach allowed us to quickly assess what is currently known about practice in the field and to update a previous US review of research (Frye et al, 2013) while using systematic review methods (e.g., Cooper, 2010). Our approach largely maintains the expected standards of a systematic review: it is rigorous, transparent, reproducible, based on explicit inclusion/exclusion criteria, and provides both quantitative and qualitative synthesis. Nevertheless, there are limitations to our approach due to the rapid timescale. In particular, we were not able to contact authors systematically to request additional information or clarification of the results in the published studies. Additionally, we did not have sufficient time to register and publish our review protocol in advance.

We judged a rigorous rapid review to be possible because a number of related reviews have recently been carried out: Frye et al.’s (2013) review for the US What Works Clearinghouse (WWC) examining the teaching of young children aged 4-8, and Simms et al.’s (2019) review of primary mathematics. This enabled us to focus our literature searches on a relatively tight timeline (2012-2019) as well as making use of the search strategies and approaches of these existing reviews. In addition, we could build on two recent secondary meta-analyses of mathematics for older pupils: Hodgen et al.’s (2018) review of mathematics teaching at Key Stages 2 and 3, and Hodgen et al.’s (2020) review of mathematics teaching for low-attaining pupils at Key Stage 3. Although these reviews focused on older pupils, the literature reviewed included studies conducted with younger children. Hence, we could ‘unpack’ the primary meta-analyses to identify additional relevant literature and examine the extent to which the general findings of these primary meta-analyses applied specifically to younger children.

Given the rapid timescale for the review, our main focus was on the effects of different interventions on attainment (rather than on attitudes or other non-cognitive outcomes).

Review outcomes

We present evidence from our review under 'strand' headings (see below), with each strand representing a key theme in the field. For each strand, our rapid review allowed us to produce two outputs:

1. A meta-analysis (where appropriate) of interventions
2. A narrative review of intervention studies

The strands – i.e. the key themes relevant to our review – were developed using an iterative approach, with the strand list refined during the data search and extraction processes. As a starting point, we drew on three sources:

1. Guidance from the Expert Panel as to areas they wished us to examine
2. The strand (or module) list developed for Hodgen, J., Foster, C., Marks, R., & Brown, M. (2018). *Evidence for Review of Mathematics Teaching: Improving Mathematics in Key Stages Two and Three: Evidence Review.* London: Education Endowment Foundation;
3. Expertise within the team in mathematics teaching and learning in Early Years and Key Stage 1.

As literature were sourced and coded, we reviewed and discussed as a team where strands could be expanded or combined or where new strands needed to be added. Strands were grouped into topic areas and intervention types. Studies could be coded to multiple strands. Interventions were also coded according to the mathematical topic addressed: Number & calculation; Shape, space & measures; Both number and geometry / spatial reasoning; Other (e.g., early algebra, such as patterning).

Data types

Our review is based on three categories of data/literature:

1. Best Evidence Syntheses, Meta-Analyses and Systematic Literature Reviews, including Expert Judgment-Based Reviews.
2. RCT (Randomised Controlled Trial) and QED (Quasi-Experimental Design) studies of interventions, published since 1/1/2012 (chosen because Frye et al, 2013, reviewed literature published up to December 2011). The full papers for studies that met our inclusion criteria were sourced for the review. Our focus on RCT and QED studies is because such studies provide robust evidence of the effects of an intervention by comparing a group receiving the intervention with a control group who do not.

3. RCT and QED studies of interventions, included in the meta-analyses above, published between 2000 and 2011, and with sufficient data to enable the extraction of an effect size that could be aggregated together with the post-2012 intervention studies using meta-analysis. These studies were only accessed indirectly through the meta-analyses, not through the original papers. This enabled us to extend the database of studies while carrying out the review within the rapid timescale.

Datasets

In order to assess the literature in a timely manner we drew on four existing substantial datasets in addition to our own searches as presented in the table below.

Publications sourced through these existing datasets were recorded within the appropriate database:

1. Best Evidence Syntheses / Systematic reviews / Meta-Analyses
2. Individual RCT and QED studies (published since 1/1/2012)
3. Individual studies (published between 2000 and 2011) where data was extracted from existing published meta-analyses

Within each database, strict notes were kept detailing the original source of the publication; i.e., the existing dataset it was identified within and/or the original meta-analysis from which an individual study was extracted.

	Existing dataset	Extent / coverage of existing dataset	How we used the dataset
1	Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide (NCEE 2014-4005).* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.	- Identified 79 studies evaluating instructional practices (with 29 meeting the What Works Clearing House stringent criteria) - Covered preschool, pre-kindergarten and kindergarten (ages 3-6) - Included studies published between 1989-2011 - Focused predominantly on US-based studies	i. Used to establish knowledge base (in line with EEF guidance for this rapid review) to the end of 2011, providing a baseline for the present review.
2	Simms, V., McKeaveney, C., Sloan, S., & Gilmore, C. (2019). *Interventions to improve mathematical achievement in primary school-aged children.* London: Nuffield Foundation.	- Systematic review of 80 primary RCT and QED studies - Covered 4-11 age-range - Included studies published between 2000-2017 - Excluded studies where children were screened to assess their need for the intervention (due, for example, to low attainment)	i. Examined authors' original complete dataset (531 studies) to identify relevant RCT/QED studies: a) Covering 3-7 age-range b) Published 2012-2017 c) Studies in non-school contexts (e.g., Early Years settings) d) Studies in which children were screened ii. Amended authors' search strings to address our RQs and extend search to April 2019
3	Hodgen, J., Foster, C., Marks, R., & Brown, M. (2018). *Evidence for Review of Mathematics Teaching: Improving Mathematics in Key Stages Two and Three: Evidence Review.* London: Education Endowment Foundation.	- Secondary meta-analysis of 66 meta-analyses and 56 other studies - Covered meta-analyses judged relevant to the 9-14 age-range (and, as a result, many of the meta-analyses included studies with younger children) - Included meta-analyses published between 1970-2017	i. Identified meta-analyses from original dataset which included 3-7 age-range (may have gone beyond, e.g. K-G8) including re-assessment of excluded meta-analyses ii. Extracted original studies from (i) which met our inclusion criteria ii. Amended authors' search strings to address our RQ and extend search to April 2019
4	Hodgen, J., Brown, M., & Coe, R. (2020). *Low attainment in mathematics: an investigation of Year 9 students*. London: Nuffield Foundation.	- Analysis of 76 meta-analyses and 31 systematic reviews - Covered meta-analyses judged relevant to low-attainers in the 11-14 age-range - Included meta-analyses published between 1970-2018 (note overlap with (2) above)	i. Identified meta-analyses from original dataset which included 3-7 age-range (may have gone beyond, e.g. K-G8) including re-assessment of excluded meta-analyses ii. Extracted original studies from (i) which met our inclusion criteria

Systematic search

In addition to drawing on the existing datasets detailed above, we also conducted our own systematic searches. There were two reasons for this:

1. To ensure search strings and results captured the agreed foci of our review, and hence triangulate and validate the existing search strategies;
2. To update existing datasets to include literature published up to April 2019.

Our searches involved four phases:

Phase 1: Updating and extending the search results of existing datasets (3) and (4) to identify further meta-analyses relevant to the present review and from which individual studies could be extracted

Phase 2: First run of systematic search for RCTs/QEDs and systematic reviews/Best Evidence Syntheses based on amended versions of search strings used in existing datasets (2) and (3).

Phase 3: Second run of systematic search for RCTs/QEDs and systematic reviews/Best Evidence Syntheses based on amended versions of search strings used in Phase 2 and updated to included emerging 'strands' and guidance from the Expert Panel.

Phase 4: Additional searches of material not picked up in Phases 1-3 including the Education Endowment Foundation (EEF) and What Works Clearing House (WWC) reports in addition to further recommendations of known reviews from the expert panel.

See Appendix 1 for more details on the search processes.

Inclusion Criteria

For RCTs and QEDs we applied the following inclusion criteria:

- Focused on mathematics (including numeracy, geometry and spatial reasoning)
- Concerned with teaching or pedagogy or strategy (i.e., exclude studies simply about children's learning)
- Has a 'well-described' intervention (or change to / deviation from teaching practice) that could be implemented (or initiated) by teachers or other adults working in Early Years or Key Stage 1 settings
- Study design involves a control group
- Conducted with children aged 3-7
- Conducted with an Early Years or Key Stage 1 setting (i.e. nursery, pre-school, reception class, school) but excluding home, child-minder and 'out of school' settings such as museums (but include nursery or school-led initiatives to help parents or carers work on maths with their child)
- Published since January 2012 ('grey' literature included)
- Not focused solely on pupils with Emotional and Behavioural Difficulties
- Written in English

Coding and data extraction

We entered all publications identified through re-analysis of existing datasets and through our searches into three separate spreadsheets representing our data categories, with a detailed record kept of the source of all citations:

1. Systematic reviews / Best Evidence Syntheses / Meta-Analyses
2. RCT (Randomised Controlled Trial) and QED (Quasi-Experimental Design) studies of interventions (including studies where data was extracted from existing meta-analyses)

Meta-analyses and Systematic reviews / Best Evidence Syntheses

We entered each meta-analysis and systematic review / BES into a spreadsheet capturing basic publication information (e.g. author(s), year of publication, title, etc.), search source and demographic information (country, ages of pupils included, area of mathematics).

From the demographic information, abstracts, and, where necessary, reading the full papers, we identified 22 of the 102 meta-analyses which addressed our review foci, although only 3 exclusively addressed the Early Years and KS1 phases.

We constructed a further spreadsheet with a separate sheet for each strand. The 22 identified meta-analyses were allocated to these strands (with some meta-analyses coded to two or more strands). One member of the team 'unpacked' each of the 22 meta-analyses and identified the underpinning studies in each which met our inclusion criteria (see inclusion criteria in systematic search phase 2). We identified 101 individual RCTs/QEDs which met our inclusion criteria from the 22 'unpacked' meta-analyses. For each of these 101 studies, we extracted the effect sizes (ES) as given in the meta-analyses (as opposed to reading the original study). Any concerns or limitations in the meta-analysis or extracted studies were additionally noted at this stage.

Another member of the team applied our inclusion criteria to the systematic review / BES spreadsheet, identifying 11 publications which met our review foci. These publications were used in the narrative on the strands only; individual studies were not extracted from these reviews.

Individual studies (RCTs and QEDs)

We entered each individual study into a spreadsheet capturing basic publication information (e.g. author(s), year of publication, title, etc.), search source and demographic information (country, ages of pupils included, area of mathematics).

The studies identified through our systematic searches were allocated to five members of the team, with 10% of the studies each being allocated to two members of the team for the purpose of inter-coder checks. Each team member checked the demographic information for their studies and identified whether the study met our inclusion criteria. Any uncertainties were flagged for discussion as a team. The team identified 96 publications from our individual searches which met the inclusion criteria. For each of these publications, full study information was extracted from the papers and detailed in the spreadsheet. Two team members checked the 10% of entries allocated for inter-coder checks for consistency of extraction and completion of the spreadsheet. Team discussions were held to address any discrepancies.

Extraction set	Details
Basic publication information	Author(s) Year of publication Title Reference Abstract Publication type
Search source	Dataset or systematic search
Strands	Identifying which topic and intervention strand(s) the study related to
Experiment type	RCT, QED, Other
Demographics	Includes intervention (yes/no)? Conducted at scale (yes/no)? Involves maths or geometry (yes/no)? Covers Key Stage 1 (5-7 years, G1)? Covers EY (PK/K)? Country study was conducted in
Effect size statistics	ES extracted from paper (if available) Number (intervention and control groups) Number of clusters (intervention and control) Low (CI-) and High (CI+) Standard error Variance Difference in Means (if pre/post not given) Pre-test mean and SD (intervention and control) Post-test mean and SD (intervention and control
Intervention name	e.g. Building Blocks, Big Maths for Little Kids
Coding notes	

Effect sizes from the 101 individual studies ‘unpacked’ from the meta-analyses were added to our individual studies spreadsheet (with ES data taken from the related meta-analysis), giving 197 studies. Of these, 25 were duplicates (i.e. they occurred in both our systematic search and in the unpacked meta-analysis studies) leaving 172 individual studies.

Meta-analysis

Of the 172 individual studies, 57 were excluded from the meta-analysis due to insufficient data to extract or calculate a comparable ES, leaving 115 studies. Of these, 14 were published pre-2012 and 102 in 2012 or later. In general, to calculate an ES we required descriptives (means and SDs) for pre- and post-test for both intervention and control groups, although we exercised judgment on studies without pre-test information.

See Appendix 2 for a flowchart providing an overview of the process.

Data analysis

As noted previously, we used the identified studies and the data extracted from them to produce two outcomes for each strand:

1. A meta-analysis (where appropriate) of interventions
2. A narrative review of intervention studies

Meta-analysis was conducted using the *R* package “metafor" (Viechtbauer, 2010), with a random-effects model, and the code used is provided in Appendix 3. The results of the meta-analyses are presented as forest plots in Appendix 5.

In producing a narrative for each strand we included some individual studies which were excluded from the meta-analysis due to insufficient information. We also drew on the, albeit limited, systematic reviews and BES examining the strand area. This was particularly important in strands where the evidence from interventions studies was limited (such as interventions to support play or executive functions).

A first draft of each strand was written by one member of the research team, then reviewed by other team members. Disagreements were resolved through discussion.

The quality (or strength) and relevance of the evidence base

Both the quality (or strength) of evidence and the relevance of the evidence for each strand were assessed using a procedure based on the GRADE system in medicine (Guyatt et al., 2008). This is an expert judgment-based approach that is informed, but not driven, by quantitative metrics (such as number of studies included).

The judgments about the quality (or strength) of evidence took account of the number of original studies, the methodological quality of the original studies, consistency of results across the studies, any reporting bias and the extent to which the findings were supported additionally by the systematic reviews and best evidence syntheses.

The judgments about relevance of the evidence and findings took account of where and when the studies were carried out, how well the interventions were defined and operationalised, any focus on particular topic areas, the age of children involved and the ease of implementation.

Members of the research team each independently gave a for each aspect, then used this to make a judgment about an overall rating. Disagreements were resolved through discussion.

See Appendix 4 for further details of this process.

References

Best Evidence Syntheses, Systematic Reviews, Expert Judgment-based Review and Existing Meta-analytic Reviews:

Anthony, G., & Walshaw, M. (2007). Effective pedagogy in mathematics/Pˆngarau: Best evidence synthesis iteration. Wellington, New Zealand: Ministry of Education.

Clements, D., Baroody, A. J., & Sarama, J. (2013). Background research on early mathematics. Background Research for the National Governor's Association (NGA) Center Project on Early Mathematics.
Retrieved from https://www.du.edu/marsicoinstitute/media/documents/dc_background_research_early_math.pdf.

Cross, C T, Woods, T A., & Schweingruber, H (Eds) [National Research Council] (2009). Mathematics Learning in Early Childhood: Paths Toward Excellence and Equity. Washington, DC: The National Academies Press. https://doi.org/10.17226/12519.

Deans for Impact (2019). The Science of Early Learning. Austin, TX: Deans for Impact.

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis. Review of Educational Research, 87(2), 243-282. doi:10.3102/0034654316687036

Dooley, T., Dunphy, E., Shiel, G., Butler, D., Corcoran, D., Farrell, T., NicMhuirí, S., O'Connor, M., & Travers, J. (2014). Mathematics in Early Childhood and Primary Education (3-8 years): Teaching and Learning (Research Report No.18). Dublin: National Council for Curriculum and Assessment *(NCCA).*

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). Teaching math to young children: A practice guide (NCEE 2014-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hodgen, J., Coe, R., Foster, C., Brown, M., Higgins, S., & Küchemann, D. (2020). *Low attainment in mathematics: an investigation focusing on Year 9 pupils in England*. London: UCL Institute of Education.

Hodgen, J., Foster, C., Marks, R., & Brown, M. (2018). *Evidence for Review of Mathematics Teaching: Improving Mathematics in Key Stages Two and Three: Evidence Review*. London: Education Endowment Foundation.

Nelson, G., & McMaster, K. L. (2019). The effects of early numeracy interventions for students in preschool and early elementary: A meta-analysis. Journal of Educational Psychology, 111(6), 1001-1022. doi:10.1037/edu0000334

Nunes, T., Bryant, P., & Watson, A. (2009). Key understandings in mathematics learning. London: Nuffield Foundation.

Pellegrini, M., Lake, C., Nietzel, A., & Slavin, R. E. (2020). *Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis*. Manuscript in preparation and provided by the authors.

Simms, V., McKeaveney, C., Sloan, S., & Gilmore, C. (2019). *Interventions to improve mathematical achievement in primary school-aged children*. London: Nuffield Foundation.

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, 27(5), 692-713.

Original studies included in the dataset and coded as 'Computer-assisted instruction, Apps and Technology Tools:

Aragón-Mendizábal, E., Aguilar-Villagrán, M., Navarro-Guzmán, J. I., & Howell, R. (2017). Improving number sense in kindergarten children with low achievement in mathematics. *Anales de Psicología, 33*(2), 311–318.

Bengtson, B. J. (2014). Changing mental representations using related physical models: The effects of analyzing number lines on learner internal scale of numerical magnitude. Dissertation Abstracts International Section A: Humanities and Social Sciences. ProQuest Information & Learning.

Bryant, B. R., Bryant, D. P., Roberts, G., & Fall, A.-M. (2016). Effects of an Early Numeracy Intervention on Struggling Kindergarteners' Mathematics Performance. International Journal for Research in Learning Disabilities, 3(1), 29–45.

Cheung, S. K., & McBride, C. (2017). Effectiveness of Parent–Child Number Board Game Playing in Promoting Chinese Kindergarteners' Numeracy Skills and Mathematics Interest. Early Education and Development, 28(5), 572–589.

Feng, S., & Caleo, J. (2000). Playing computer games versus better learning. *Paper presented at the Annual Conference of the Eastern Educational Research Association*. Clearwater, FL

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.

Fletcher, N. (2016). Development and evaluation of a computer program to teach symmetry to young children. Dissertation Abstracts International Section A: Humanities and Social Sciences. ProQuest Information & Learning.

Foster, M. E., Anthony, J. L., Clements, D. H., Sarama, J., & Williams, J. J. (2018). Hispanic dual language learning kindergarten students' response to a numeracy intervention: A randomized control trial. Early Childhood Research Quarterly, 43, 83-95.

Foster, M. E., Anthony, J. L., Clements, D. H., Sarama, J., & Williams, J. M. (2016). Improving mathematics learning of kindergarten students through computer-assisted instruction. Journal for Research in Mathematics Education, 47(3), 206-232.

Fuchs, L. S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., & Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. Journal of Learning Disabilities, 39(5), 467–475. doi: 10.1177/00222194060390050701

González-Castro, P., Cueli, M., Cabeza, L., & Rodríguez, C. (2014). Improving basic math skills through integrated dynamic representation strategies. Psicothema, 26(3), 378-384.

Jabagchourian, J. J. (2008). Incremental changes in children's multi-digit number representations and arithmetic procedures: Linking strategies and concepts in early mathematics (Doctoral dissertation). Available from ProQuest Dissertations & Theses database. (UMI No. 3319844)

Jamalian, A. (2014). Grouping Gestures Promote Children's Effective Counting Strategies by Adding a Layer of Meaning Through Action (Doctoral dissertation, Teachers College).

Kermani, H. (2017). Computer Mathematics Games and Conditions for Enhancing Young Children's Learning of Number Sense. Malaysian Journal of Learning and Instruction, 14(2), 23–57.

Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., & Reynvoet, B. (2016). Enhancing arithmetic in pre-schoolers with comparison or number line estimation training: Does it matter?. Learning and Instruction, 46, 1-11.

Magnolia Consulting (2012). A final report for the evaluation of Pearson's Waterford Early Learning Program: Year 2. Magnolia Consulting

Mohd Syah, N. E., Hamzaid, N. A., Murphy, B. P., & Lim, E. (2016). Development of computer play pedagogy intervention for children with low conceptual understanding in basic mathematics operation using the dyscalculia feature approach. Interactive Learning Environments, 24(7), 1477-1496.

Nemmi, F., Helander, E., Helenius, O., Almeida, R., Hassler, M., Räsänen, P., & Klingberg, T. (2016). Behavior and neuroimaging at baseline predict individual response to combined mathematical and working memory training in children. Developmental Cognitive Neuroscience, 20, 43-51.

Nunes, T. (2019) *An evaluation of the onebillion maths apps using an RCT design*. London: The Education Endowment Foundation

Nunes, T., Barros, R., Evangelou, M., Strand, S., Mathers, S. & Sanders-Ellis, D. (2018), *1stClass@Number Evaluation report and executive summary*, London: The Education Endowment Foundation.

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction, 23*, 125-135.

Outhwaite, L. A., Gulliford, A., Pitchford, N. J., & Faulder, M. (2019). Raising Early Achievement in Math With Interactive Apps: A Randomized Control Trial. Journal of Educational Psychology, 111(2), 284–298.

Pagar, D. L. (2013). The effects of a grouping by tens manipulative on children's strategy use, base ten understanding and mathematical knowledge (Doctoral dissertation, Teachers College).

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.

Pitchford, N. J. (2015). Development of early mathematical skills with a tablet intervention: A randomized control trial in Malawi. Frontiers in Psychology, 6.

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(1), 56–65.

Salminen, J. B., Koponen, T. K., Leskinen, M., Poikkeus, A. M., & Aro, M. T. (2015). Individual variance in responsiveness to early computerized mathematics intervention. Learning and Individual Differences, 43, 124–131. doi: 10.1016/j.lindif.2015.09.002

Sarama, J., & Clements, D. H. (2015). Scaling up early mathematics interventions: Transitioning with trajectories and technologies. In Mathematics and transition to school (pp. 153-169). Springer, Singapore.

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, 74–88. doi: 10.1080/10409289.2015.1057462

Shamir, A., & Baruch, D. (2012). Educational e-books: A support for vocabulary and early math for children at risk for learning disabilities. Educational Media International, 49(1), 33-47. http://dx.doi.org/10.1080/09523987.2012.662623

Tournaki, N., Bae, Y. S., & Kerekes, J. (2008). Rekenrek: A manipulative used to teach addition and subtraction to students with learning disabilities. Learning Disabilities: A Contemporary Journal, 6(2), 41-59.

Uzomah, S. L. (2012, January 1). Teaching Mathematics to Kindergarten Students through a Multisensory Approach. ProQuest LLC. ProQuest LLC.

Wang, H., & Woodworth, K. (2011). Evaluation of Rocketship Education's use of DreamBox Learning's online mathematics program. Menlo Park, CA: SRI International.

Wilson, A.J., Dehaene, S., Dubois, O., & fayol, M. (2009) Efects of an adaptive game intervention on accessing number sense in low-socioeconomic status kindergarten children. *Mind Brain and Education, 3* (4), 224-234

Zaranis, N. (2016). The use of ICT in kindergarten for teaching addition based on realistic mathematics education. *Education and Information Technologies, 21*(3), 589–606.

Zaranis, N. (2018). Comparing the Effectiveness of Using ICT for Teaching Geometrical Shapes in Kindergarten and the First Grade. *International Journal of Web-Based Learning and Teaching Technologies, 13*(1), 50–63.

Original studies included in the dataset and coded as 'Technological tools':

Sung, W., Ahn, J., & Black, J. B. (2017). Introducing Computational Thinking to Young Learners: Practicing Computational Perspectives through Embodiment in Mathematics Education. *Technology, Knowledge and Learning, 22*(3), 443–463.

Original studies included in the dataset and coded as 'Cooperative learning':

Meloni, C., Fanari, R., Bertucci, A., & Berretti, S. (2017). Impact of Early Numeracy Training on Kindergarteners from Middle-Income Families. International Association for Development of the Information Society, 14.

Tarim, K. (2009). The effects of cooperative learning on preschoolers' mathematics problem-solving ability. Educational Studies in Mathematics, 72(3), 325–340. doi:10.1007/s10649-009-9197-x

Original studies included in the dataset and coded as 'Peer tutoring':

Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners'mathematical development: Effects of peer-assisted learning strategies. The Elementary School Journal, 101 (5), 495–509. doi:10.1086/499684

Fuchs, L. S., Fuchs, D., Yazdian, L., & Powell, S. R. (2002). Enhancing first-grade children's mathematical development with peer-assisted learning strategies. School Psychology Review, 31(4), 569-83.

Original studies included in the dataset and coded as 'executive functions or metacognition':

Baten, E., Praet, M., & Desoete, A. (2017). The relevance and efficacy of metacognition for instructional design in the domain of mathematics. *ZDM Mathematics Education, 49,* 613-623. https://doi.org/10.1007/s11858-017-0851-y

Kroesbergen, E. H., van't Noordende, J. E., & Kolkman, M. E. (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Child*

Neuropsychology, 20, 23-37.

Nemmi, F., Helander, E., Helenius, O., Almeida, R., Hassler, M., Räsänen, P., & Klingberg, T. (2016). Behavior and neuroimaging at baseline predict individual response to combined mathematical and working memory training in children. *Developmental Cognitive Neuroscience, 20,* 43-51.

Rittle-Johnson, B., Saylor, M., & Swygert, K. E. (2008). Learning from explaining: Does it matter if mom is listening? *Journal of Experimental Child Psychology, 100*, 215–224.

Original studies included in the dataset and coded as 'explicit teaching':

Bryant, B. R., Bryant, D. P., Roberts, G., & Fall, A.-M. (2016). Effects of an Early Numeracy Intervention on Struggling Kindergarteners' Mathematics Performance. International Journal for Research in Learning Disabilities, 3(1), 29–45.

Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., et al. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. Exceptional Children, 78(1), 7–23. doi:10.1177/001440291107800101

Cheng, Z. J. (2012). Teaching young children decomposition strategies to solve addition problems: An experimental study. The Journal of Mathematical Behavior, 31(1), 29-47.

Clarke, B., Baker, S., Smolkowski, K., Doabler, C., Strand Cary, M., & Fien, H. (2015). Investigating the Efficacy of a Core Kindergarten Mathematics Curriculum to Improve Student Mathematics Learning Outcomes. Journal of Research on Educational Effectiveness, 8(3), 303–324.

Clarke, B., Doabler, C. T., Kosty, D., Kurtz Nelson, E., Smolkowski, K., Fien, H., & Turtura, J. (2017). Testing the Efficacy of a Kindergarten Mathematics Intervention by Small Group Size. AERA Open, 3(2).

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Cary, M. S. (2014). Examining the efficacy of a tier 2 kindergarten mathematics intervention. Journal of Learning Disabilities, 49, 152–165. doi: 10.1177/0022219414538514

Clarke, B., Doabler, C. T., Strand Cary, M., Kosty, D., Baker, S., Fien, H., & Smolkowski, K. (2014). Preliminary evaluation of a tier 2 mathematics intervention for first-grade students: using a theory of change to guide formative evaluation activities. School Psychology Review, 43, 160–177

Clarke, B., Doabler, C., Smolkowski, K., Kurtz Nelson, E., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a Tier 2 kindergarten mathematics intervention. Journal of Research on Educational Effectiveness, 9(4), 607-634.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the building blocks project. Journal for Research in Mathematics Education, 38(2), 136-163

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. American Educational Research Journal, 45(2), 443-494

Codding, R. S., Chan-Iannetta, L., George, S., Ferreira, K., & Volpe, R. (2011). Early number skills: Examining the effects of class-wide interventions on kindergarten performance. School Psychology Quarterly, 26, 85–96. doi: 10.1037/a0022661

Davies, R. S., & Qudisat, R. M. (2015). Benefits of the Michiana Daily Mathtracks Programme for students living in poverty. Educational Research and Evaluation, 21(4), 301-323.

Doabler, C. T., Clarke, B., Kosty, D. B., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2016). Testing the efficacy of a tier 2 mathematics intervention: A conceptual replication study. Exceptional Children, 83(1), 92-110.

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. Journal of Learning Disabilities, 46(2), 166–181. doi:10.1177/0022219411410233

Dyson, N., Jordan, N. C., Beliakoff, A., & Hassinger-Das, B. (2015). A kindergarten number-sense intervention with contrasting practice conditions for low-achieving children. Journal for research in mathematics education, 46(3), 331.

Foster, M. E., Anthony, J. L., Clements, D. H., Sarama, J., & Williams, J. J. (2018). Hispanic dual language learning kindergarten students' response to a numeracy intervention: A randomized control trial. Early Childhood Research Quarterly, 43, 83-95.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. Journal of Educational Psychology, 97, 493–513. doi: 10.1037/0022-0663.97.3.493

Gersten, R., Rolfhus, E., Clarke, B., Decker, L. E., Wilkins, C., & Dimino, J. (2015). Intervention for first graders with limited number knowledge: Large-scale replication of a randomized controlled trial. American Educational Research Journal, 52(3), 516-546.

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building kindergarteners’ number sense: A randomized controlled study. Journal of Educational Psychology, 104(3), 647–660. doi:10.1037/a0029018

Kaufmann, L., Delazer, M., Pohl, R., Semenza, C., & Dowker, A. (2005). Effects of aspecific numeracy educational program in kindergarten children: A pilot study. Educational Research and Evaluation, 5, 405–431. doi: 10.1080/13803610500110497

Kermani, H. (2017). Computer Mathematics Games and Conditions for Enhancing Young Children’s Learning of Number Sense. Malaysian Journal of Learning and Instruction, 14(2), 23–57.

McKenzie, M. A., Marchand-Martella, N. E., Moors, M. E., & Martella, R. C. (2004). Teaching basic math skills to preschoolers using Connecting Math Concepts Level K. Journal of Direct Instruction, 4(1), 85-94. [185]

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2017). Improving mathematics teaching in kindergarten with realistic mathematical education. Early Childhood Education Journal, 45(3), 369–378.

Passolunghi, M. C., & Costa, H. M. (2016). Working memory and early numeracy training in preschool children. Child Neuropsychology, 22, 81–98. doi: 10.1080/09297049.2014.971726

Purpura, D. J., Baroody, A. J., Eiland, M. D., Reid, E. E., & Society for Research on Educational Effectiveness (SREE). (2012). Fostering First-Graders’ Reasoning Strategies with the Most Basic Sums. Society for Research on Educational Effectiveness. Society for Research on Educational Effectiveness.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves lowincome preschoolers’ numerical understanding. Journal of Educational Psychology, 101, 545−560. doi:10.1037/a0014239

Stockard, J. & Engelmann, K. (2010). Academic kindergarten and later academic success: The impact of Direct Instruction. Journal of Behavioral Assessment and Intervention for Children, 1(1), 2-24.

Toll, S. W., & Van Luit, J. E. H. (2013). Accelerating the early numeracy development of kindergartners with limited working memory skills through remedial education. Research in Developmental Disabilities, 34, 745–755. doi: 10.1016/j.ridd.2012.09.003

Original studies included in the dataset and coded as ‘feedback and formative assessment’:

Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., et al. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children*, 78(1), 7–23. doi:10.1177/001440291107800101

Fuchs, L. S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., & Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities*, 39(5), 467–475. doi: 10.1177/00222194060390050701

Fuchs, L. S., Compton, D. L., Fuchs, D., Hamlett, C. L., DeSelms, J., Seethaler, P. M., et al. (2013). Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology*, 105(1), 58–77. doi:10.1037/a0030127

Polly, D., Wang, C., Martin, C., Lambert, R. G., Pugalee, D. K., & Middleton, C. W. (2017). The Influence of an Internet-Based Formative Assessment Tool on Primary Grades Students’ Number Sense Achievement. *School Science and Mathematics*, 117(3), 127–136.

Popa, N. L., & Pauc, R. L. (2015). Dynamic Assessment, Potential Giftedness and Mathematics Achievement in Elementary School. *Acta Didactica Napocensia*, 8(2), 23–32.

Sarama, J., & Clements, D. H. (2015). Scaling up early mathematics interventions: Transitioning with trajectories and technologies. In *Mathematics and transition to school* (pp. 153-169). Springer, Singapore.

Original studies included in the dataset and coded as ‘manipulatives and representations’:

Booth, J.L., & Siegler, R.S. (2008). Numerical magnitude representations influence arithmetic learning. Child Development, 79, 4, pp1016 -1031

Casey, B. M., Andrews, N., Schindler, H, Kersh, J. E., Samper, A., & Copley, J. (2008). The development of spatial skills through interventions involving block building activities. Cognition and Instruction, 26, 269-309.

Elofsson, J., Gustafson, S., Samuelsson, J., & Träff, U. (2016). Playing number board games supports 5-year-old children’s early mathematical development. The Journal of Mathematical Behavior, 43, 134-147.

González-Castro, P., Cueli, M., Cabeza, L., & Rodríguez, C. (2014). Improving basic math skills through integrated dynamic representation strategies. Psicothema, 26(3), 378-384.

Jamalian, A. (2014). Grouping Gestures Promote Children's Effective Counting Strategies by Adding a Layer of Meaning Through Action (Doctoral dissertation, Teachers College).

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building Kindergartners’ Number Sense: A Randomized Controlled Study. Journal of Educational Psychology, 104(3), 647–660. https://doi.org/10.1037/a0029018

Kidd, J. K., Curby, T. W., Boyer, C. E., Gadzichowski, K. M., Gallington, D. A., Machado, J. A., & Pasnak, R. (2012). Benefits of an intervention focused on oddity and seriation. Early Education & Development, 23(6), 900-918.

Kidd, J., Pasnak, R., Gadzichowski, K. M., Gallington, D., McKnight, P., Boyer, C., & Carlson, A. (2014). Instructing First-Grade Children on Patterning Improves Reading and Mathematics. Early Education and Development, 25(1), 134–151.

Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., & Reynvoet, B. (2016). Enhancing arithmetic in pre-schoolers with comparison or number line estimation training: Does it matter?. Learning and Instruction, 46, 1-11.

Mohd Syah, N. E., Hamzaid, N. A., Murphy, B. P., & Lim, E. (2016). Development of computer play pedagogy intervention for children with low conceptual understanding in basic mathematics operation using the dyscalculia feature approach. Interactive Learning Environments, 24(7), 1477-1496.

Nemmi, F., Helander, E., Helenius, O., Almeida, R., Hassler, M., Räsänen, P., & Klingberg, T. (2016). Behavior and neuroimaging at baseline predict individual response to combined mathematical and working memory training in children. Developmental Cognitive Neuroscience, 20, 43-51.

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. Learning and Instruction, 23, 125-135.

Pagar, D. L. (2013). The effects of a grouping by tens manipulative on children's strategy use, base ten understanding and mathematical knowledge (Doctoral dissertation, Teachers College).

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.

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.

Ruiter, M., Loyens, S., & Paas, F. (2015). Watch your step children! Learning two-digit numbers through mirror-based observation of self-initiated body movements. Educational Psychology Review, 27(3), 457-474.

Uzomah, S. L. (2012, January 1). Teaching Mathematics to Kindergarten Students through a Multisensory Approach. ProQuest LLC. ProQuest LLC.

Original studies included in the dataset and coded as 'movement':

Have, M., Nielsen, J. H., Ernst, M. T., Gejl, A. K., Fredens, K., Grøntved, A., & Kristensen, P. L. (2018). Classroom-based physical activity improves children's math achievement – A randomized controlled trial. PLoS ONE, 13(12).

Jamalian, A. (2014). Grouping Gestures Promote Children's Effective Counting Strategies by Adding a Layer of Meaning Through Action (Doctoral dissertation, Teachers College).

Ruiter, M., Loyens, S., & Paas, F. (2015). Watch your step children! Learning two-digit numbers through mirror-based observation of self-initiated body movements. Educational Psychology Review, 27(3), 457-474.

Shoval, E., Sharir, T., Arnon, M., & Tenenbaum, G. (2018). The Effect of Integrating Movement into the Learning Environment of Kindergarten Children on their Academic Achievements. Early Childhood Education Journal, 46(3), 355-364.

Sung, W., Ahn, J., & Black, J. B. (2017). Introducing Computational Thinking to Young Learners: Practicing Computational Perspectives through Embodiment in Mathematics Education. Technology, Knowledge and Learning, 22(3), 443–463.

Original studies included in the dataset and coded as 'play':

Shirani Bidabadi, Nasr Esfahani, Mirshah Jafari, & Abedi. (2019). Developing a mathematics curriculum to improve learning behaviors and mathematics competency of children. *The Journal of Educational Research, 112*(3), 421-428. doi:10.1080/00220671.2018.1547960.

Vogt, F., Hauser, B., Stebler, R., Rechsteiner, K., & Urech, C. (2018). Learning through Play--

Pedagogy and Learning Outcomes in Early Childhood Mathematics. European Early Childhood Education Research Journal, 26(4), 589–603. doi: 10.1080/1350293X.2018.1487160

Original studies included in the dataset and coded as 'storybooks':

Casey, B., Erkut, S., Ceder, I., & Young, J.M. (2008). Use of a storytelling context to improve girls' and boys' geometry skills in kindergarten. Journal of Applied Developmental Psychology, 29, 1, pp29-48

Hassinger-Das, B., Jordan, N. C., & Dyson, N. (2015). Reading stories to learn math: Mathematics vocabulary instruction for children with early numeracy difficulties. Elementary School Journal, 116(2), 242–264.

Purpura, D. J., Napoli, A. R., Wehrspann, E. A., & Gold, Z. S. (2017). Causal connections between mathematical language and mathematical knowledge: A dialogic reading Intervention. Journal of Research on Educational Effectiveness, 10(1), 116-137.

Segal-Drori, O., Kalmanovich, L. B. H., & Shamir, A. (2018). Electronic book for promoting emergent math: A comparison between kindergarteners at risk for learning disabilities and with typical development. Journal of Educational Computing Research. DOI: https://doi.org/10.1177/0735633118769459

Shamir, A. & Baruch, D. (2012). Educational e-books: A support for vocabulary and early math for children at risk for learning disabilities. Educational Media International, 49(1), 33-47.

Van den Heuvel-Panhuizen, M., Elia, I., & Robitzsch, A. (2016). Effects of reading picture books on kindergartners' mathematics performance. Educational Psychology, 36(2), 323-346.

Original studies included in the dataset and coded as 'mathematical talk':

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.

[*] Russo, J., & Hopkins, S. (2018). Teaching primary mathematics with challenging tasks: How should lessons be structured?. *The Journal of Educational Research*, 1-12.

Original studies included in the dataset and coded as 'tutoring by adults':

Bryant, B. R., Bryant, D. P., Roberts, G., & Fall, A.-M. (2016). Effects of an Early Numeracy Intervention on Struggling Kindergarteners' Mathematics Performance. International *Journal for Research in Learning Disabilities, 3*(1), 29–45.

Clarke, B., Doabler, C. T., Kosty, D., Kurtz Nelson, E., Smolkowski, K., Fien, H., & Turtura, J. (2017). Testing the Efficacy of a Kindergarten Mathematics Intervention by Small Group Size. *AERA Open, 3*(2).

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Cary, M. S. (2014). Examining the efficacy of a tier 2 kindergarten mathematics intervention. Journal of Learning Disabilities, 49, 152–165. doi: 10.1177/0022219414538514

Clarke, B., Doabler, C. T., Strand Cary, M., Kosty, D., Baker, S., Fien, H., & Smolkowski, K. (2014). Preliminary Evaluation of a Tier 2 Mathematics Intervention for First-Grade Students: Using a Theory of Change to Guide Formative Evaluation Activities. School Psychology Review, 43(2), 160-177.

Clarke, B., Doabler, C., Smolkowski, K., Nelson, E. K., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a tier 2 kindergarten mathematics intervention. Journal of Research on Educational Effectiveness, 9, 607–634. doi: 10.1080/19345747.2015.1116034

Doabler, C. T., Clarke, B., Kosty, D. B., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2016). Testing the efficacy of a tier 2 mathematics intervention: A conceptual replication study. Exceptional Children, 83, 92–110. doi: 10.1177/0014402916660084

Doabler, C. T., Clarke, B., Stoolmiller, M., Kosty, D. B., Fien, H., Smolkowski, K., & Baker, S. K. (2017). Explicit Instructional Interactions: Exploring the Black Box of a Tier 2 Mathematics Intervention. Remedial and Special Education, 38(2), 98-110.

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. Journal of Learning Disabilities, 46(2), 166–181.

Dyson, N., Jordan, N. C., Beliakoff, A., & Hassinger-Das, B. (2015). A kindergarten number-sense intervention with contrasting practice conditions for low-achieving children. Journal for research in mathematics education, 46(3), 331.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C., Hamlett, C. L., ... Bryant, J. D. (2013). Effects of first-grade number knowledge tutoring with contrasting forms of practice. Journal of Educational Psychology, 105, 58–77. doi: 10.1037/a0030127

Gersten, R., Rolfhus, E., Clarke, B., Decker, L. E., Wilkins, C., & Dimino, J. (2015). Intervention for first graders with limited number knowledge: Large-scale replication of a randomized controlled trial. American Educational Research Journal, 52(3), 516-546.

Hassinger-Das, B., Jordan, N. C., & Dyson, N. (2015). Reading stories to learn math: Mathematics vocabulary instruction for children with early numeracy difficulties. Elementary School Journal, 116(2), 242–264.

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.

Smith, T. M., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating Math Recovery: Assessing the Causal Impact of a Diagnostic Tutoring Program on Student Achievement. American Educational Research Journal, 50(2), 397–428. https://doi.org/10.3102/0002831212469045

Torgerson, C. J., Wiggins, A., Torgerson, D., Ainsworth, H., & Hewitt, C. (2013). Every Child Counts: Testing policy effectiveness using a randomised controlled trial, designed, conducted and reported to CONSORT standards. Research In Mathematics Education, 15(2), 141-153.

Original studies included in the dataset and coded as 'whole curriculum interventions':

Alger, M. W. (2015). Evaluating Early Numeracy Skills in Preschool Children: A Program Evaluation of Rural Head Start Classrooms (Doctoral dissertation, Alfred University, Alfred, NY).

Casa, T. M., Firmender, J. M., Gavin, M. K., & Carroll, S. R. (2017). Kindergarteners' Achievement on Geometry and Measurement Units That Incorporate a Gifted Education Approach. Gifted Child Quarterly, 61(1), 52–72.

Clarke, B., Baker, S., Smolkowski, K., Doabler, C., Strand Cary, M., & Fien, H. (2015). Investigating the Efficacy of a Core Kindergarten Mathematics Curriculum to Improve Student Mathematics Learning Outcomes. Journal of Research on Educational Effectiveness, 8(3), 303–324.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum:

Summative research on the building blocks project. Journal for Research in Mathematics Education, 38(2), 136-163

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. American Educational Research Journal, 45(2), 443-494

Gavin, M. K., Casa, T. M., Firmender, J. M., & Carroll, S. R. (2013). The Impact of Advanced Geometry and Measurement Curriculum Units on the Mathematics Achievement of First-Grade Students. Gifted Child Quarterly, 57(2), 71–84.

Khomais, S. F. (2014). Enhancing preschool children's number knowledge: the suitability of an intervention programme for Saudi practice. Early Child Development and Care, 184(1), 32-49.

Kinzie, M. B., Whittaker, J. V., Williford, A. P., DeCoster, J., McGuire, P., Lee, Y., & Kilday, C. R. (2014). MyTeachingPartner-Math/Science pre-kindergarten curricula and teacher supports: Associations with children’s mathematics and science learning. Early Childhood Research Quarterly, 29(4), 586–599. https://doi.org/10.1016/j.ecresq.2014.06.007

Lewis Presser, A., Clements, M., Ginsburg, H., & Ertle, B. (2015). Big math for little kids: The effectiveness of a preschool and kindergarten mathematics curriculum. Early education and development, 26(3), 399-426.

Llorente, C., Pasnik, S., Moorthy, S., Hupert, N., Rosenfeld, D., & Gerard, S. (2015). Preschool Teachers Can Use a PBS KIDS Transmedia Curriculum Supplement to Support Young Children's Mathematics Learning: Results of a Randomized Controlled Trial. Society for Research on Educational Effectiveness.

Mattera, S. K., Jacob, R., Morris, P. A., & MDRC. (2018). Strengthening Children’s Math Skills with Enhanced Instruction: The Impacts of Making Pre-K Count and High 5s on Kindergarten Outcomes. MDRC. MDRC.

Sarama, J., & Clements, D. H. (2015). Scaling up early mathematics interventions: Transitioning with trajectories and technologies. In Mathematics and transition to school (pp. 153-169). Springer, Singapore.

Stokes, L., Hudson-Sharp, N., Dorsett, R., Rolfe, H., Anders, J., George, A., Buzzeo, J. & Munro-Lott, N. (2018) Mathematical Reasoning: Evaluation report and executive summary. London: Education Endowment Foundation

Worth, J., Sizmur, J., Ager, R., & Styles, B. (2015). Improving Numeracy and Literacy: Evaluation Report and Executive Summary. Education Endowment Foundation.

Original studies included in the dataset and coded as ‘professional development’:

Robinson-Smith, L., Fairhurst, C., Stone, G., Bell, K., Elliott, L., Gascoine, L., ... & Torgerson, D. (2018). Maths Champions Evaluation Report and Executive Summary. London: The Education Endowment Foundation.

Other references:

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25. doi:10.1080/0969594x.2010.513678

Blair, & Razza. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev, 78*(2), 647-663. doi:10.1111/j.1467-8624.2007.01019.x

Blatchford, P., Bassett, P., Brown, P., & Webster, R. (2009). The effect of support staff on pupil engagement and individual attention. *British Educational Research Journal, 35*(5), 661-686. Retrieved from http://www.informaworld.com/10.1080/01411920902878917

Brooks, G., Phal, K., Pollard, A., & Rees, F. (2008). Effective and inclusive practices in family literacy, language and numeracy: A review of programmes and practice in the UK and internationally. London: NRDC.

Cara, O., & Brooks, G. (2012). Evidence of the Wider Benefits of Family Learning: A Scoping Review. BIS Research Paper Number 93. London: Department for Business, Innovation and Skills.

Cheung, S. K., & McBride, C. (2017). Effectiveness of Parent–Child Number Board Game Playing in Promoting Chinese Kindergarteners' Numeracy Skills and Mathematics Interest. Early Education and Development, 28(5), 572–589.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic Press.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the building blocks project. Journal for Research in Mathematics Education, 38(2), 136-163

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. American Educational Research Journal, 45(2), 443–494. doi:10.3102/0002831207312908

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. Journal for Research in Mathematics Education, 42(2), 127–166.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. Journal for Research in Mathematics Education, 42(2), 127–166.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. Journal for Research in Mathematics Education, 42(2), 127–166.

Colliver, Y. (2018). Fostering young children's interest in numeracy through demonstration of its value: the Footsteps Study. Mathematics Education Research Journal, 30(4), 407–428.

Cooper, H. (2010). *Research synthesis and meta-analysis*. Thousand Oaks, CA: Sage.

Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., . . . Coe, R. (2015). Developing Great Teaching: Lessons from the international reviews into effective professional development. London: Teacher Development Trust.

Gascoine, L., Higgins, S., & Wall, K. (2017). The assessment of metacognition in children aged 4–16 years: a systematic review. *Review of Education, 5*(1), 3-57. doi:10.1002/rev3.3077

Goldin-Meadow, S. (2011). Learning through gesture. *WIREs Cognitive Science, 2*(6), 595-607. doi:10.1002/wcs.132

Gueudet, G., Bosch, M., diSessa, A., Kwon, O.-N., & Verschaffel, L. (2016). Transitions in mathematics education (ICME-13 Topical Surveys). New York, NY: Springer.

Hembree, R. (1992). Experiments and Relational Studies in Problem Solving: A Meta-Analysis. Journal for Research in Mathematics Education, 23(3), 242-273. doi:10.2307/749120

Hodgen, J., Adkins, M., Ainsworth, S., & Evans, S. (2019). Catch Up® Numeracy: Evaluation report and executive summary. London: Education Endowment Foundation.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence. Review of Educational Research, 88(4), 547-588. doi:10.3102/0034654318759268

McKenzie, M. A., Marchand-Martella, N. E., Moore, M. E., & Martella, R. C. (2004). Teaching basic math skills to preschoolers using Connecting Math Concepts Level K. *Journal of Direct Instruction, 4*(1), 85-94.

Melby-Lervag, M., Redick, T., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of 'far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science, 11*, 512-534.

Millett, A., Brown, M., & Askew, M. (Eds.). (2004). *Primary mathematics and the developing professional*. Dordrecht: Kluwer.

Muijs, D., & Bokhove, C. (2020). *Metacognition and Self-regulation: Evidence Review*. London: Education Endowment Foundation.

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. doi:https://doi.org/10.1016/j.jecp.2017.10.002

Rutt, S., Easton, C., & Stacey, O. (2014). Catch Up Numeracy: evaluation report and executive summary. London: Education Endowment Foundation.Sharples, J., Webster, R., & Blatchford, P. (2015). Making best use of teaching assistants: Guidance Report. London: Education Endowment Foundation.

Sarama, J., & Clements, D. H. (2009). Building blocks and cognitive building blocks: Playing to know the world mathematically. *American Journal of Play, 1*(3), 313-337.

Slavin, R. E., & Lake, C. (2008). Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis. *Review of Educational Research, 78*(3), 427-515. doi:10.3102/0034654308317473

Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: the mediating role of teachers' zones of enactment. *Journal of Curriculum Studies, 31*(2), 143 - 175.

Stockard, J. & Engelmann, K. (2010). Academic kindergarten and later academic success: The impact of Direct Instruction. Journal of Behavioral Assessment and Intervention for Children, 1(1), 2-24. [315]

van den Heuvel-Panhuizen, M., & Elia, I. (2012). Developing a framework for the evaluation of picturebooks that support kindergartners' learning of mathematics. Research in Mathematics Education, 14(1), 17-47. doi:10.1080/14794802.2012.657437

Verschaffel, L., Torbeyns, J., & De Smedt, B. (2017). Young children's early mathematical competencies: Analysis and stimulation. In T. Dooley & G. Gueudet (Eds.), Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education (CERME10, February 1-5, 2017) (pp. 31-52). Dublin, Ireland: DCU Institute of Education and ERME. http://www.mathematik.uni-dortmund.de/ieem/erme_temp/CERME10_Proceedings_final.pdf

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.

Warhurst, C., Nickson, D., Commander, J., & Gilbert, K. (2013). 'Role stretch': assessing the blurring of teaching and non-teaching in the classroom assistant role in Scotland. British Educational Research Journal, 40(1), 170-186. doi:10.1002/berj.3036

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided Play: Principles and Practices. *Current Directions in Psychological Science, 25*(3), 177-182. doi:10.1177/0963721416645512

Williams, P. (2008). *Independent Review of Mathematics Teaching in Early Years Settings and Primary Schools*. Nottingham: DCSF.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Appendix 1: Details of the Systematic Searches

The systematic search was carried out in four phases as follows.

Phase 1

This phase of the systematic search focused on updating our meta-analyses database. We took the search strings used in Dataset 3 (see Hodgen, Foster, Marks, & Brown, 2018, Appendix 15 pp.153-199). From these, we:

- Identified and excluded all search terms/strings irrelevant to the present review (e.g. those focussed on strands or age-phases sitting outside of the foci of our present review);
- Removed all search strings developed to identify "research review" OR "research synthesis" OR "review of research" (due to focus on meta-analytic studies);
- Added in new search terms to address the strand foci of the present review.

From this we produced the following search terms. All permutations were combined in developing our search strings.

General	Specific		Literature Type
	Original search terms	New search terms	
mathematic* math* numeracy arithmetic education pedagogy intervention* strateg* teach* learn* instruction	concrete apparatus diagram* imagery manipulative mastery professional development resource* textbook* transition visualization*	co-operative Learning direct instruction explicit instruction feedback group-work heuristic* metacognition parent* play self-instruction self-regulation student centred tutor*	a meta-analysis a meta-analytic meta-analysis meta-analytic quantitative synthesis

For search strings including *original* specific search terms (column 2) we used a date range of March 2017 to April 2019 to account for this search building on previous datasets. For search strings including *new* specific search terms (column 3) we used a date range of 1970 to April 2019 to account for this search adding to previous datasets.

All search strings were run as full text searches across multiple databases and search engines:

- ArticleFirst OCLC
- British Education Index
- Child Development & Adolescent Studies
- ECO
- Education Abstracts
- EducatiOnline
- ERIC
- Erikson Early Math Collaborative
- Google / Google Scholar

- JSTOR
- MathSciNet via EBSCOhost
- PapersFirst OCLC
- PEDAL
- ProQuest
- PsycARTICLES
- PsycINFO
- Teacher Reference Center

Additional hand-searches were also made of review journals from March 2017 – April 2019:

- Educational Research
- Educational Research Review
- Educational Researcher
- Review of Educational Research
- Review of Research in Education
- Review of Education
- Open Review of Educational Research

Inclusion criteria for meta-analyses were developed from the inclusion/exclusion criteria used within dataset (3) in addition to the specific inclusion criteria covering other phases of this systematic review:

- Focused on mathematics (including numeracy, geometry and spatial reasoning)
- Concerned with teaching or pedagogy or strategy (i.e. exclude studies simply about children's learning)
- Included children aged 3-7 (possibly as part of a broader review)
- Included / relevant to an Early Years or Key Stage 1 setting (i.e. nursery, pre-school, reception class, school) but excluding home, child-minder and 'out of school' settings such as museums (but include nursery or school-led initiatives to help parents or carers work on maths with their child)
- Published since 1970 / 2017 (see above) re. *original* and *new* specific search terms
- Not focussed solely on pupils with EBD (e.g. Losinski, M. L., Ennis, R. P., Sanders, S. A., & Nelson, J. A. (2019). A Meta-Analysis Examining the Evidence-Base of Mathematical Interventions for Students With Emotional Disturbances. *The Journal of Special Education, 52*(4), 228-241.)
- Written in English

We used these inclusion criteria to assess the citations amassed from the searches. We assessed citations on the basis of their titles wherever possible and on abstracts where required. Papers were not accessed or read at this stage.

Phase 1 of our systematic review produced the following additions to our database:

Search string	Search dates	Hits	New extracted on title (i.e. not in existing database)	Remaining after initial cleaning / removal of duplicates
Search strings including *original* specific search terms [Databases, Google Scholar and hand journal searches]	March 2017 – April 2019	1038	54	56
Search strings including *new* specific search terms [Databases, Google Scholar and hand journal searches]	1970 – April 2019	2180	17	

Systematic search: Phase 2

Phase 2 of the systematic search focused on updating our RCT and QED individual studies database, updating our Systematic Review database and on capturing any relevant meta-analyses not picked up on Phase 1.

We took the search string used in Dataset 2 (see Simms, McKeaveney, Sloan, & Gilmore, 2019, p.9). From this, we:

- Re-ran the Dataset 2 search string in its original form with the date range of January 2017 – April 2019 in order to capture studies published post the search date of the original dataset
- Amended the 'population' element of the Dataset 2 search string to focus on the population of our present review, namely Early Years and Key Stage One, with the date range of January 2012 to April 2019 to capture studies published post Dataset 1 (the baseline for our present review)

We also took the search string used in the original Dataset 3 (as in Phase 1, but including all review search terms), and:

- Amended the 'population' element of the Dataset 3 search string to focus on the population of our present review, namely Early Years and Key Stage One
- Included all *original* and *new* specific search terms (as detailed in Phase 1) and added further specific search terms (talk, discussion and outdoor exploration) following discussion with the Expert Panel/EEF

From the above, we produced three separate search strings:

Source	Date range of search	Search string
Original Dataset 2 search string	January 2017 to April 2019	(Primary OR Elementary OR Kindergarten* OR “Grade 1” OR “Grade 2”, “Grade 3”, “Grade 4”, “Grade 5”) AND (school* OR educat* OR class* OR teach* OR learn* OR instruct* OR train* OR program*) AND (Math* OR “Number Sense” OR Numer* OR Arithmetic* OR counting OR addition OR subtraction OR multiplication OR division OR Adding OR Geometry OR fractions OR algebra OR "place value") AND (Achieve* OR “Standardi* Test” OR Anxiety OR Attitud* OR “Self-Efficacy” OR Confidence OR Enjoyment) AND (Trial OR RCT OR Quasi OR Random* OR “Control Group” OR “Post Test” OR experimental)
Dataset 2 search string with amended population	January 2012 to April 2019	(“Early Years” OR Reception OR Nursery OR “Foundation Stage”) AND (school* OR educat* OR class* OR teach* OR learn* OR instruct* OR train* OR program*) AND (Math* OR “Number Sense” OR Numer* OR Arithmetic* OR counting OR addition OR subtraction OR multiplication OR division OR Adding OR Geometry OR fractions OR algebra OR "place value") AND (Achieve* OR “Standardi* Test” OR Anxiety OR Attitud* OR “Self-Efficacy” OR Confidence OR Enjoyment) AND (Trial OR RCT OR Quasi OR Random* OR “Control Group” OR “Post Test” OR experimental)
Dataset 3 search string with amended population and foci	January 2012 to April 2019	(“Early Years” OR Reception OR Nursery OR “Foundation Stage” OR Elementary OR Kindergarten* OR “Grade 1” OR “Year 1” OR “Year 2” OR “Key Stage One”) AND (mathematic* OR math* OR numeracy OR arithmetic OR education OR pedagogy OR intervention* OR strateg* OR teach* OR learn* OR instruction) AND (manipulative OR “concrete apparatus” OR imagery OR visualization* OR diagram* OR textbook* OR resource* OR “professional development” OR parent* OR play OR mastery OR transition OR “explicit instruction” OR “direct instruction” OR tutor* OR heuristic* OR feedback OR “self-instruction” OR “metacognition” OR “self-regulation” OR “co-operative learning” OR “group work” OR “student centred” OR “outdoor exploration” OR talk OR discussion) AND (“a meta-analysis” OR “a meta-analytic” OR “meta-analysis” OR “meta-analytic” OR “quantitative synthesis” OR “best evidence synthesis” OR “systematic review” OR “research review” OR “research synthesis” OR “review of research”)

Each search string was run for full text searches across multiple databases and search engine as detailed in Phase 1. Additional hand-searches were also made of relevant journals in mathematics education and early childhood education / development from March 2017 to April 2019.

For located meta-analyses and syntheses, the same inclusion criteria / process as in Phase 1 was employed. For RCTs and QEDs we applied the following inclusion criteria:

- Focused on mathematics (including numeracy, geometry and spatial reasoning)
- Concerned with teaching or pedagogy or strategy (i.e. exclude studies simply about children’s learning)
- Has a ‘well-described’ intervention (or change to / deviation from teaching practice) that could be implemented (or initiated) by teachers or other adults working in Early Years or Key Stage 1 settings
- Study design involves a control group
- Conducted with children aged 3-7
- Conducted with / relevant to an Early Years or Key Stage 1 setting (i.e. nursery, pre-school, reception class, school) but excluding home, child-minder and ‘out of school’ settings such as museums (but include nursery or school-led initiatives to help parents or carers work on maths with their child)
- Published since January 2012
- Not focussed solely on pupils with EBD
- Written in English

We used these inclusion criteria to assess the citations amassed from the searches. We assessed citations on the basis of their titles wherever possible and on abstracts where required. Papers were not accessed or read at this stage.

Phase 2 of our systematic review produced the following additions to our databases:

Search string	Search dates	Hits	New extracted on title (i.e. not in existing database)	Remaining after initial cleaning / removal of duplicates
Original Dataset 2 search string	January 2017 to April 2019	2928	82	134
Dataset 2 search string with amended population	January 2012 to April 2019	15,883	35	
Dataset 3 search string with amended population and foci	January 2012 to April 2019	607	58	

Systematic search: Phase 3

Phase 3 of the systematic search focused on further updating our RCT and QED individual studies database, updating our Systematic Review database and on capturing any relevant meta-analyses not picked up on Phase 1. Phase 3 was identical to Phase 2 but with amendments to the specific search strings to reflect our growing awareness of the themes / strands relevant to the field and further guidance from the Expert Panel. Hence, the search strings used were:

Source	Date range of search	Search string
Dataset 2 search string with amended population and foci	January 2012 to April 2019	(“Early Years” OR Reception OR Nursery OR “Foundation Stage” OR Elementary OR Kindergarten* OR “Grade 1” OR “Year 1” OR “Year 2” OR “Key Stage One”) AND (school* OR educat* OR class* OR teach* OR learn* OR instruct* OR train* OR program*) AND (games OR construction OR pretend OR “role-play” OR “role play” OR co-construction OR calculation OR conservation OR shape OR measures OR “data handling” OR statistics OR “problem-solving” OR “problem solving” OR communication OR reasoning OR connections OR misconceptions OR “working memory” OR “mathematical objects” OR “number line” OR pictorial OR visual OR technology OR CAI OR “computer aided instruction” OR coding OR programming OR “programable toy” OR tablets OR integrated OR “cross-curricular” OR “story time” OR “picture books” OR rhymes OR songs OR “language development” OR vocabulary OR assessment OR “progress monitoring” OR intervention OR “task choice” OR “pedagogical content knowledge” OR “teacher knowledge” OR “teaching assistant” OR “TA” OR “home learning” OR “home environment” OR “head-start” OR “head start” OR surestart OR “sure start” OR “early intervention” OR “Dynamic assessment”) AND (Achieve* OR “Standardi* Test” OR Anxiety OR Attitud* OR “Self-Efficacy” OR Confidence OR Enjoyment) AND (Trial OR RCT OR Quasi OR Random* OR “Control Group” OR “Post Test” OR experimental)
Dataset 3 search string with further amended population and foci	January 2012 to April 2019	(“Early Years” OR Reception OR Nursery OR “Foundation Stage” OR Elementary OR Kindergarten* OR “Grade 1” OR “Year 1” OR “Year 2” OR “Key Stage One”) AND (mathematic* OR math* OR numeracy OR arithmetic OR education OR pedagogy OR intervention* OR strateg* OR teach* OR learn* OR instruction) AND (games OR construction OR pretend OR “role-play” OR “role play” OR co-construction OR calculation OR conservation OR shape OR measures OR “data handling” OR statistics OR “problem-solving” OR “problem solving” OR communication OR reasoning OR connections OR misconceptions OR “working memory” OR “mathematical objects” OR “number line” OR pictorial OR visual OR technology OR CAI OR “computer aided instruction” OR coding OR programming OR “programable toy” OR tablets OR integrated OR “cross-curricular” OR “story time” OR “picture books” OR rhymes OR songs OR “language development” OR vocabulary OR assessment OR “progress monitoring” OR intervention OR “task choice” OR “pedagogical content knowledge” OR “teacher knowledge” OR “teaching assistant” OR “TA” OR “home learning” OR “home environment” OR “head-start” OR “head start” OR surestart OR “sure start” OR “early intervention” OR “Dynamic assessment”) AND (“a meta-analysis” OR “a meta-analytic” OR “meta-analysis” OR “meta-analytic” OR “quantitative synthesis” OR “best evidence synthesis” OR “systematic review” OR “research review” OR “research synthesis” OR “review of research”)

The search was then run identically to that described in Phase 2 but with the date-range on both searches running from January 2012 to April 2019.

Phase 3 of our systematic review produced the following additions to our databases:

Search string	Search dates	Hits	New extracted on title (i.e. not in existing database)	Remaining after initial cleaning / removal of duplicates
Amended Dataset 2 search string with amended population and foci	January 2012 to April 2019	7553	197	131
Dataset 3 search string with further amended population and foci	January 2012 to April 2019	504	39	

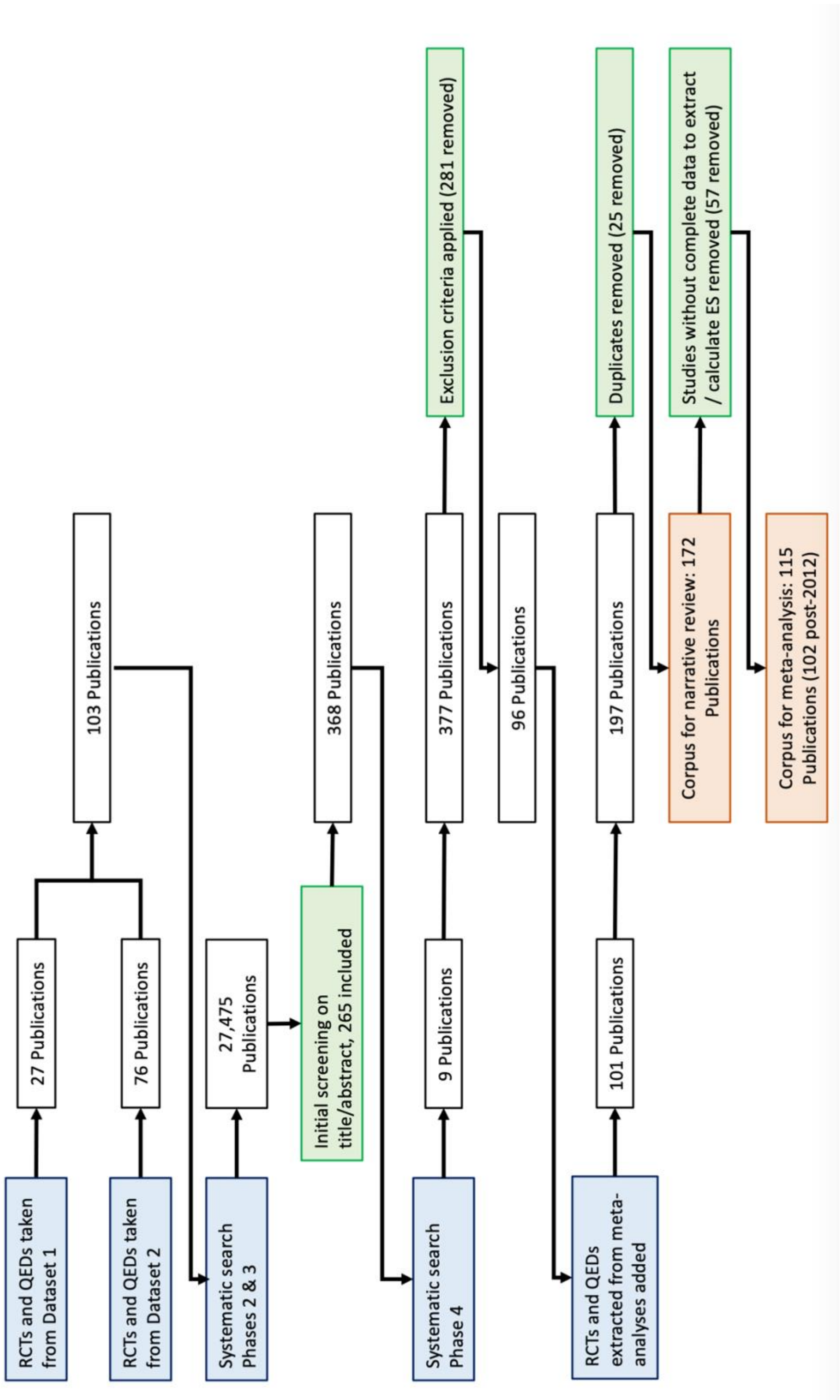
Systematic search: Phase 4

In order to ensure our searches were comprehensive, we identified in advance a set of studies that should be identified and included in the review. Across Phases 1 to 3 we noted that searches were not identifying several of these studies. These included reports and reviews published by the Education Endowment Foundation (EEF) and the What Works Clearing House (WCC). We established that these were not identified as they are not included in the key academic literature databases. As a result, one member of the research team completed the following:

Search source	Process and outcomes
Hand-searched WWC Website: *Find What Works:* [*https://ies.ed.gov/ncee/wwc/*](https://ies.ed.gov/ncee/wwc/)	• Filtered by Pre-K and Mathematics & K-12 & Mathematics. Identified studies (dated 2011 or after AND reviewed 2013 or after) AND conducted with (Pre-K, K or Grade 1) • Excluded studies focused on literacy and oracy with no maths intervention, but where maths outcomes were reported, but did include The Creative Curriculum for Preschool, Fourth Edition (physical play and sand & water) • Excluded 'global' teacher education programmes e.g., Teach for America (K-12) and TAP: The System for Teacher and Student Advancement • Some repeated studies found (e.g. Scott Foresman-Addison Wesley Elementary Mathematics already covered by Saxon Maths study) • Included related reviews where relevant (although few WWC webpages identified related studies) • Searched for original (key) publications where available
Searched for Best Evidence Syntheses on Ministry of Education and related websites in Ireland, New Zealand and Scotland on the basis of advice from EEF EY & KS1 Expert Guidance Panel.	• This process identified 3 potentially relevant reports from CfBT and CREC
Hand-searched EEF Website: *Completed Projects:* [*https://educationendowmentfoundation.org.uk/projects-and-evaluation/reports/*](https://educationendowmentfoundation.org.uk/projects-and-evaluation/reports/)	• Filtered completed projects by mathematics and by Early Years and KS1: This identified 6 potential project • Analysed literature underpinning the EEF Early Years toolkit for meta-analyses and single studies relevant to numeracy

In total, Phase 4 of our Systematic Search added 3 meta-analyses, 3 systematic reviews/reports and 9 single studies to our databases.

Appendix 2: Flowchart of process of identifying RCTS and QEDs



Appendix 3: R code used for meta-analyses

```
install.packages("metafor")
library(metafor)

Data <- read.csv("Metadata.csv", header = TRUE, sep=",")
Data <- Data[!(Data$Exclude=="Y"),]
Data <- subset(Data, manipulatives=="Y")
Data$effectsize <- as.numeric(as.character( Data$effectsize))
Data$var <- as.numeric(as.character( Data$var))

res <- rma(yi=effectsize, vi=var, data=Data, slab=paste(author, year, sep=", "),
method="REML")
res
forest(res, xlab="Effect size (d)")
mtext(bquote(paste("Manipulatives and representations: (Q = ",
    .(formatC(res$QE, digits=2, format="f")), ", df = ", .(res$k - res$p),
    ", p = ", .(formatC(res$QEp, digits=2, format="f")), "; ", I^2, " = ",
    .(formatC(res$I2, digits=1, format="f")), "%)")))
```

Appendix 4: Judging quality (or strength) and relevance of evidence

Quality (or strength) of evidence

This tables details how the review team made judgments about the quality of the body of evidence about a specific type of intervention or strategy, and the extent to which the findings are supported by a robust body of evidence. The term ‘quality’ is used in preference to ‘strength’ to avoid confusion with the size of effects. Each member of the review team made independent judgments, which were then compared, aggregated and moderated. Disagreements were to be discussed as a team, although, in the event, this was not necessary.

Aspects of strength of evidence	**Grade [0, minimal, to 3, strong]**	**Notes**
A: The number of original studies		• Thresholds: 20 studies, strong [3] (e.g., CAI, explicit teaching); 5 or less, low [1]; and none as minimal [0] • For strong grade, at least 2 studies conducted at scale (>500 pupils in study & > 250 in intervention group)
B: The methodological quality of the original studies		• Sample size? Well-reported? Design appropriate (including clustered analysis) • Ideally, we would ask whether aspects of implementation considered but I do not think we have captured this as yet.
C: Consistency of results across the studies		• Are the results consistent across studies and is the intervention sufficiently similar (and coherently described) across the studies? Are any differences sufficiently well explained?
D: Any reporting bias		• Is there any indication (or evidence) of publication bias?
E: Evidence from systematic reviews and BES		• Do the reviews support the findings of the original studies? If not, are there good reasons for the differences?
Overall judgment of the strength of evidence		Make overall judgment based on above criteria, then moderate across the team.

Relevance to English Early Years and Key Stage 1 mathematics teaching and learning

This table details how the review team made judgments about the relevance of the specific type of intervention or strategy to Early Years and Key Stage 1 mathematics contexts in England. Each member of the review team made independent judgments, which were then compared, aggregated and moderated. Disagreements were to be discussed as a team, although, in the event, this was not necessary. Relevance is not independent of the quality of the body of evidence, so the overall relevance grading cannot be more than one grade higher than than the quality of evidence grading.

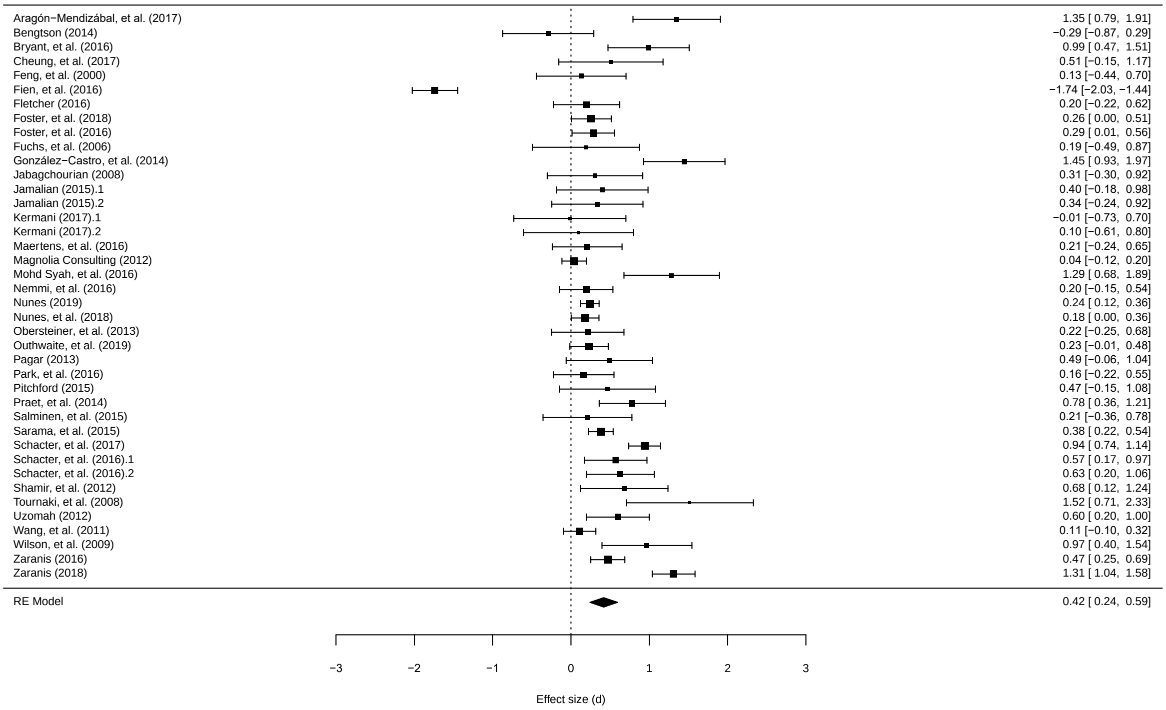
Aspects of relevance	**Grade [0, minimal, to 3, high]**	**Notes**
A: Where and when the studies were carried out		• Were any (a large proportion of?) studies carried out in England? Should we have thresholds for this (which would need to be low to be operationally useful in discriminating between strands)? • Were the studies carried out in educational systems or contexts judged to be similar to England (either similar overall or similar for the topic)? • If mostly US, is this aspect of US mathematics education judged to be sufficiently similar to England to be relevant? • If many of the studies are dated, is this a threat relevance?
B: How the interventions were defined and operationalised		• Are the interventions either available in England or sufficiently well-described for teachers to implement in England? • Are there widely available examples of use in England?
C: Any focus on particular topic areas		• Are the studies skewed towards particular mathematical topics – both broad topics (number/calculation v shape/space/geometry v measures) and more specific (narrow) topics?
D: Age of children /phase of education		• Were the studies carried out across the age range – and in different kinds of context (more / less formal)? • Are there reasons why the intervention is more appropriate for either EY or KS1?
E: Ease of implementation??		• Are there potential difficulties with implementation (e.g., cost, amount of training required, level of external support required)?
Overall relevance judgment		Make overall judgment based on above criteria, then moderate across the team. Focus more attention on criteria A and B with C, D and E as caveats.

Appendix 5: Results of meta-analyses

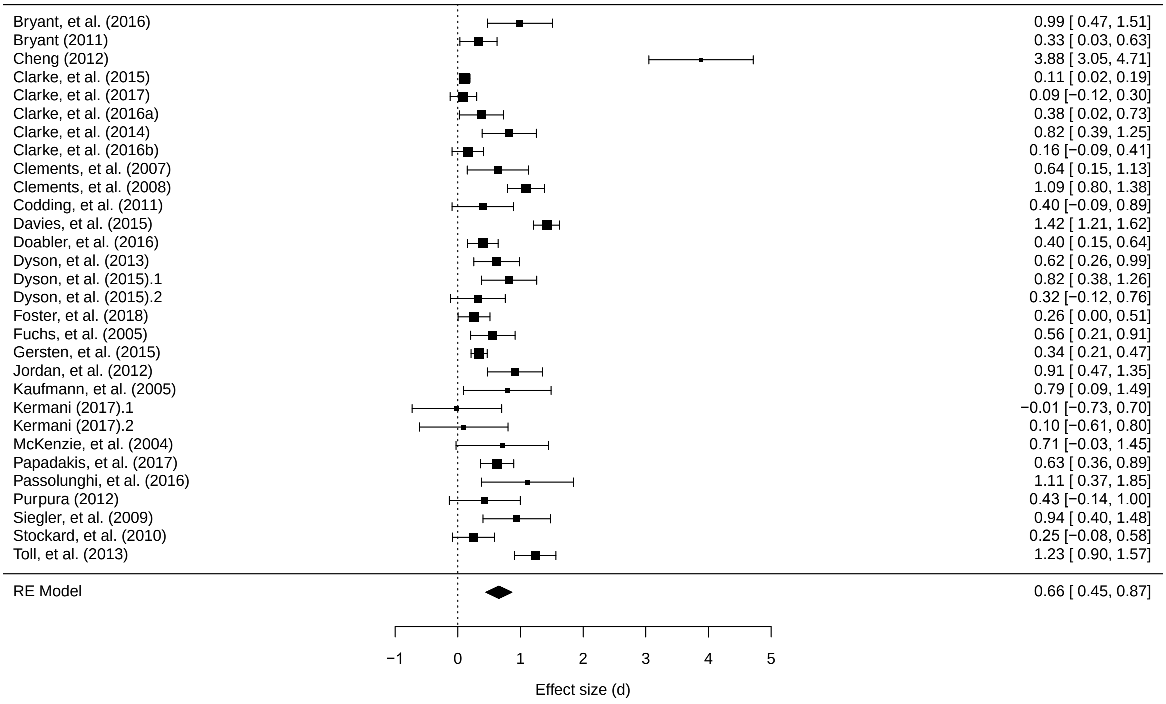
The results of meta-analyses for the strands where this was judged appropriate are summarised in the table below. The detailed results for each intervention are presented here as forest plots.

Strand / Intervention	**Aggregated Effect Size (or impact on attainment)**		**95% CI**	**I^2**
Computer-assisted instruction and apps	0.42	Moderate	(0.24, 0.59)	91.3%
Explicit teaching	0.66	Moderate	(0.45, 0.87)	93.3%
Individual and small-group tutoring by adults	0.50	Moderate	(0.37, 0.64)	77.8%
Manipulatives and representations	0.34	Moderate	(0.07, 0.60)	90.6%
Whole-curriculum interventions	0.44	Moderate	(0.16, 0.72)	97.9%
Feedback and formative assessment	0.31	Moderate	(0.18, 0.44)	59.7%
Use of storybooks (reported with mathematical talk)	0.96	Large	(0.29, 1.63)	90.7%

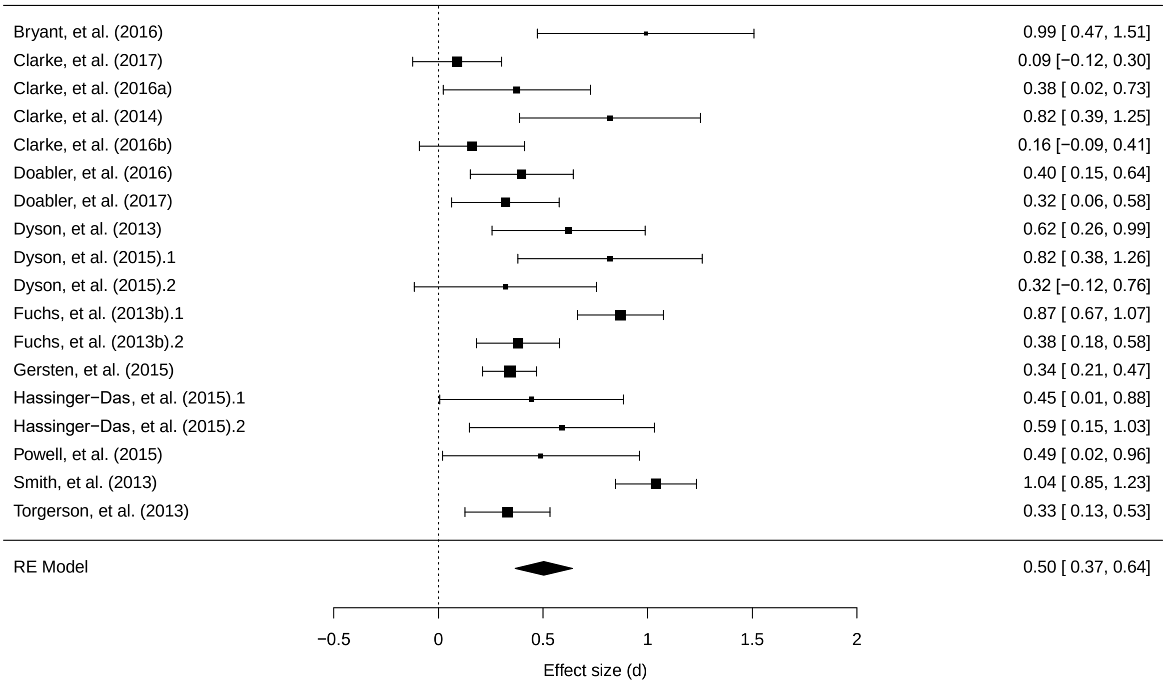
Computer−aided instruction, apps and technology tools: (Q = 383.31, df = 39, p = 0.00; I^2 = 91.3%)



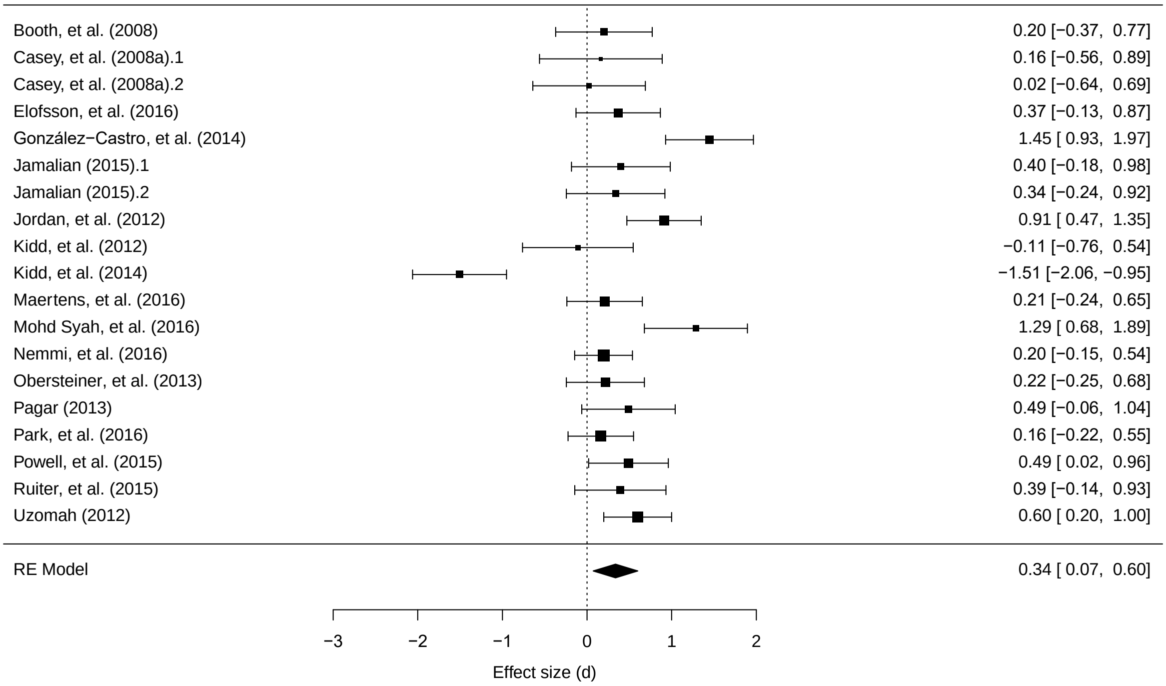
Explicit teaching: (Q = 300.35, df = 29, p = 0.00; I^2 = 93.3%)



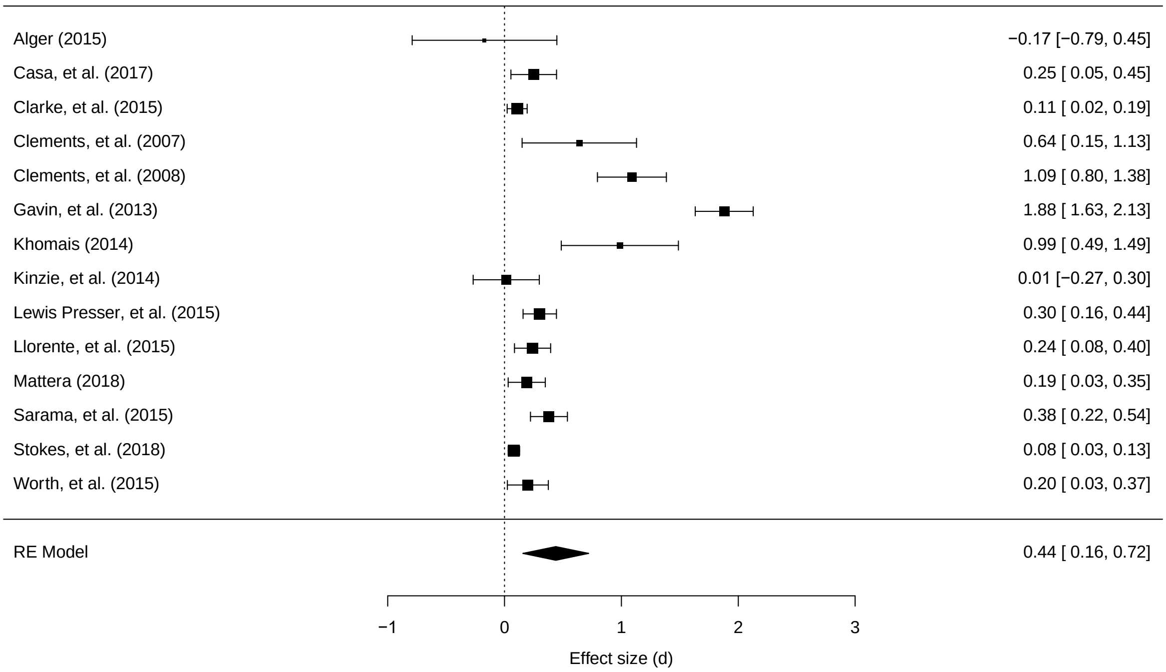
Individual and small−group tutor ing by adults: (Q = 84.75, df = 17, p = 0.00; I^2 = 77.8%)



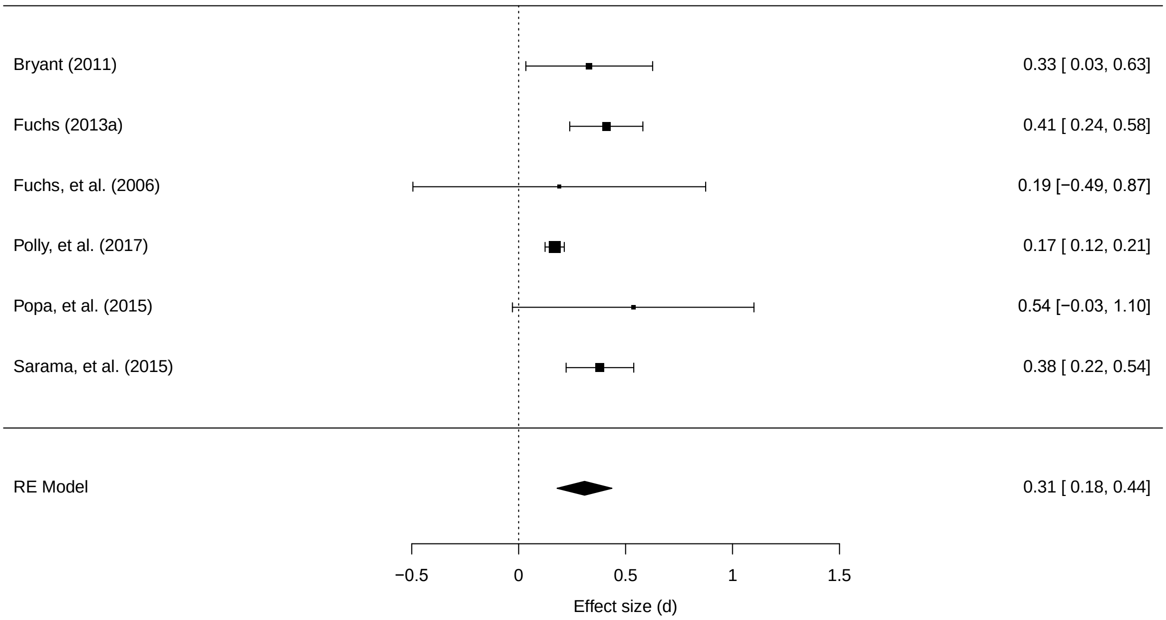
Manipulatives and representations: (Q = 83.32, df = 18, p = 0.00; I^2 = 80.6%)

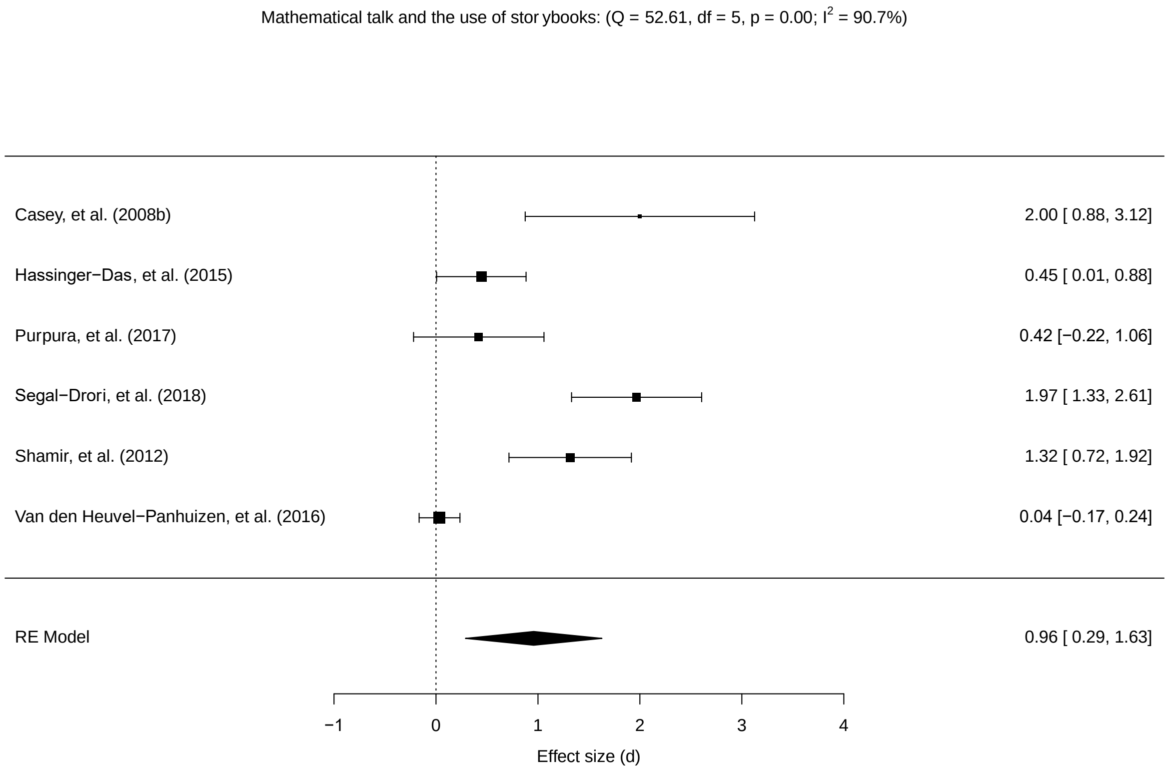


Whole−curriculum interventions: (Q = 261.02, df = 13, p = 0.00; I^2 = 97.9%)



Feedback and formative assessment: (Q = 14.74, df = 5, p = 0.01; I^2 = 59.7%)





Mathematical talk and the use of stor ybooks: (Q = 52.61, df = 5, p = 0.00; I^2 = 90.7%)

Appendix 6: Guidelines for selecting computer software

Cross et al. (2009, p.253) suggest that the following criteria should be considered when selecting software:

- Actions and graphics should provide a meaningful context for children.
- Reading level, assumed attention span, and way of responding should be appropriate for the age level. Instructions should be clear, such as simple choices in the form of a picture menu.
- After initial adult support, children should be able to use the software independently. There should be multiple opportunities for success.
- Feedback should be informative.
- Children should be in control. Software should provide as much manipulative power as possible.
- Software should allow children to create, program, or invent new activities. It should have the potential for independent use but should also challenge. It should be flexible and allow more than one correct response.