Education Endowment Foundation

Mathematics Mastery

Secondary Evaluation Report

February 2015

Independent evaluators:

John Jerrim
Helen Austerberry
Cosette Crisan
Anne Ingold
Candia Morgan
Dave Pratt
Cathy Smith
Meg Wiggins

IOE LONDON

Leading education and social research
Institute of Education
University of London

The Education Endowment Foundation (EEF)

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- Identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- Evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale;
- Encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust, as lead charity in partnership with Impetus Trust (now part of Impetus-The Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Robbie Coleman
Research and Communications Manager
Education Endowment Foundation
9th Floor, Millbank Tower
21-24 Millbank
SW1P 4QP

p: 020 7802 1679
e: robbie.coleman@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

The project was managed by John Jerrim. Dave Pratt led the process evaluation, with the assistance of Cosette Crisan, Candia Morgan, and Cathy Smith. Helen Austerberry, Anne Ingold, and Meg Wiggins played key roles in contacting and liaising with schools during testing. Tests were supplied by an external contractor (GL Assessment). Staff from the Ark children's charity (Michael Mann, Laura Bastyan, and Helen Drury) led the intervention within schools.

Contact details:

Dr John Jerrim
Institute of Education, University of London
20 Bedford Way
London
WC1H 0AL

p: 07590 761 755
e: j.jerrim@ioe.ac.uk

Contents

Abbreviations

APS	**Average point score**
CI	**Confidence interval**
EAL	**English as an additional language**
FSM	**Free School Meals**
ICC	**Inter-cluster correlation**
IoE	**Institute of Education**
ITT	**Intention-to-treat**
KS1	**Key Stage 1**
KS2	**Key Stage 2**
MICE	**Multiple Imputation by Chained Equations**
MM	**Mathematics Mastery**
NPD	**National Pupil Database**
OLS	**Ordinary Least Squares**
PiM	**Progress in Maths**
RCT	**Randomised controlled trial**
UPN	**Unique Pupil Number**
URN	**Unique Record Number**

Executive summary

The project

The Mathematics Mastery programme is a whole-school approach to teaching mathematics that aims to raise attainment for all pupils and close the attainment gap between pupils from low income families and their peers. The programme aims to deepen pupils' conceptual understanding of key mathematical concepts. Compared to traditional curricula, fewer topics are covered in more depth, and greater emphasis is placed on problem solving and on encouraging mathematical thinking.

This evaluation assessed the impact of Mathematics Mastery on pupils in Year 7, after the programme had been implemented in schools for one year. It was intended that schools would also begin to use the programme in Year 8 in the second year of implementation, and continue until the approach was in place across the school. 44 schools from London and the South East participated in the trial, with a total sample of 5,938 pupils. Participating schools received training and resources to support the adoption of the programme, which was delivered by the education charity Ark.

The project was one of two evaluations of Mathematics Mastery funded by the Education Endowment Foundation (EEF). A second project assessed the impact of Mathematics Mastery on pupils in Year 1. An overall summary combining findings from both evaluations is available on the EEF website.

Key conclusions
1. On average, Year 7 pupils in schools adopting Mathematics Mastery made a small amount more progress than pupils in schools that did not. However, the effect detected was not statistically significant, meaning that it is not possible to rule out chance as an explanation.
2. There is no strong evidence that the approach had a greater impact on lower-attaining pupils than on higher-attaining pupils.
3. Combining the findings from this study and a second randomised controlled trial of Mathematics Mastery involving Year 1 pupils may strengthen the overall evidence for the approach.
4. Given the low per-pupil cost, Mathematics Mastery may represent a cost-effective change for schools to consider. However, teachers would need to resolve tensions related to differentiation to provide support for all groups of children.
5. It would be worthwhile to track the medium- and long-term impact of the approach, to assess whether there is a cumulative effect to the approach and whether it has an impact on performance in high-stakes tests.

What impact did it have?

On average, pupils in schools adopting Mathematics Mastery made more progress than similar pupils in schools that did not adopt the programme. The small positive effect can be estimated as equivalent to approximately one month's additional progress. However, the effect was not statistically significant, meaning that it is not possible to determine that it did not occur by chance. A similar average impact was found for pupils eligible for free school meals.

There is no strong evidence that the approach had a greater impact on lower-attaining pupils than on higher-attaining pupils. Pupils made more progress in material that was focused on within the programme. Despite Mathematics Mastery not covering calculator use within its Year 7 syllabus, no negative side-effects were observed for this aspect of children's maths skills. Possible explanations for the small average effects include the relatively little exposure children would have had to the programme and the fact that this was the first year the programme had been introduced. There were wide variations in how schools responded to the intervention. Nevertheless, there was evidence of a shift in most schools away from the teaching of procedures towards a problem-solving approach, involving increased use of discussion, objects, and diagrams.

Some tensions arose where teachers felt unsure about the impact of the intervention on examination results, in particular in relation to the seemingly reduced content coverage in each year. In addition, some teachers were unsure how to differentiate between different ability groups when tasks were open and sometimes did not have explicit learning objectives. In a follow-up study, GCSE results will be used to evaluate the long-term impact of the programme.

How secure is this finding?

Overall, the findings from this evaluation are judged to be of moderate to high security. The evaluation was set up as an effectiveness trial, meaning that it aimed to test the programme under realistic conditions in a large number of schools. The evaluation used a randomised controlled trial design, with schools randomly allocated to adopt the programme or continue with 'business as usual'. Randomisation reduced the likelihood that there were unobservable differences between schools in each group, and increased the security of the findings.

To help assess whether the improvement should be attributed to the programme, it is possible to combine the findings from this trial with other evaluations of Mathematics Mastery. This approach, known as a 'meta-analysis', can lead to a more accurate estimate of an intervention's effect. However, it is also important to note the limitations of meta-analysis, and the care needed in interpreting findings based on studies that may vary in important ways. Combining the findings from this study and a second randomised controlled trial of Mathematics Mastery involving Year 1 pupils shows a statistically significant average impact of one additional month's progress. This combined finding strengthens the evidence for the approach overall, and is discussed in further depth in a summary report on the EEF's website.

Of the schools that initially enrolled in the trial, 88% of schools and 73% of pupils who initially enrolled in the trial were successfully followed through to completion. Participating schools volunteered to take part, so it is not possible to say whether similar effects would be seen in all schools. Participating schools had lower-achieving, lower-income and more ethnic minority pupils than the country as a whole. The extent to which the low-stakes test results used in this evaluation are predicative of high-stakes exams is difficult to assess at this stage.

How much does it cost?

The cost of the approach is estimated to be approximately £7,460 in the first year for a secondary school, including teacher training costs. The average 'per pupil' cost of the intervention is therefore around £50 per year, in the first year, with cost per pupil likely to reduce in future years.

Group	No. of pupils (schools)	Effect size (95% confidence interval)	Estimated months' progress	Evidence strength*	Cost
All pupils vs. comparison	**5,938 pupils (44 schools)**	**+0.06 (−0.04 to +0.15)**	**+1 month**	🔒🔒🔒🔒 (4 of 5)	**£**
FSM pupils vs. comparison	**1,610 pupils (44 schools)**	**+0.07 (−0.04 to +0.17)**	**+1 month**		**£**

*For more information about evidence ratings, see Appendix 1 in the main evaluation report.
Evidence ratings are not provided for sub-group analyses, which will always be less secure than overall findings

Introduction

Intervention

The Mathematics Mastery programme is a whole-school approach to teaching mathematics that aims to raise attainment for all pupils and close the attainment gap between pupils from low income families and their peers. The programme aims to deepen pupils' conceptual understanding of key mathematical concepts. Compared to traditional curricula, fewer topics are covered in more depth and greater emphasis is placed on problem solving and on encouraging mathematical thinking.

This clustered Randomised Controlled Trial (RCT) investigated whether implementing the Mathematics Mastery programme led to improvement in Year 7 pupils' maths test scores.

Background evidence

A number of meta-analyses have been conducted on the effectiveness of 'mastery' approaches to teaching and learning, with several examples included in the Sutton Trust-EEF Toolkit.[1,2]

Typically, mastery approaches involve breaking down subject matter and learning content into discrete units with clear objectives and pursuing these objectives until they are achieved before moving on to the next unit. Students are generally required to show high levels of achievement before progressing to master new content. This approach differs from conventional approaches, which often cover a specified curriculum at a particular pre-determined pace.

Synthesising evidence from 46 studies, Guskey and Piggott (1988) suggest mastery learning is associated with a 0.60 standard deviation increase in pupils' learning outcomes. Kulik, Kulik, and Bangert-Drowns (1990) review 108 studies, and find an average effect size of 0.52, though with more pronounced effects for weaker students. Similarly, Waxman et al. (1985) examine the effect of adaptive education on a range of cognitive and behavioural outcomes, reporting an average effect size of 0.45 across 38 studies. Thus, existing evidence does suggest that such 'mastery' approaches to teaching may have a positive effect.

It is important to note that the Mathematics Mastery programme differs from some examples of mastery learning previously studied. For example, a key feature of many apparently effective programmes studied to date was that once pupils have completed each block of content they must demonstrate a high level of success on a test, typically at about the 80% level. Pupils not meeting this target would receive additional instruction, while those who succeeded would engage in enrichment activity that sought to deepen their understanding of the same topic. This is a different approach to that adopted by the Mathematics Mastery programme, within which the developers sought to provide all pupils with "opportunities to deepen understanding through enrichment throughout their time studying the content". In Mathematics Mastery, the class spent longer than usual on each concept or procedure the first time they studied it, but they did not significantly delay the starting of new topics in the event that some pupils were still unable to achieve mastery. Rather, the intention was that the majority would achieve a good understanding of the key ideas in the required time, and that intervention would be provided for any pupils at risk of falling behind.

In addition, there are a number of other limitations or caveats related to the existing evidence base. First, the meta-analyses reviewed above are 25 to 30 years old, with many of the studies reviewed conducted in the 1960s, 1970s, and early 1980s. Classroom environments, teaching quality, pedagogy, and social context may have changed considerably since. Second, some studies included

---

[1] See http://educationendowmentfoundation.org.uk/toolkit/mastery-learning/.

[2] See http://educationendowmentfoundation.org.uk/toolkit/mastery-learning/references-mastery-learning.

in the mastery section of the toolkit show small or no effects, suggesting that making mastery learning work effectively in all circumstances is challenging. Third, most of the studies contained within these meta-analyses took place in the United States. This is a specific context and the results may not generalise. Fourth, the evidence is not specifically about mastery techniques applied to the learning of maths, and so may not necessarily be indicative of the impact of the Mathematics Mastery programme. Hence, the relevance of such evidence for contemporary education policy in England (and Western countries more generally) may be limited.

Evaluation objectives

The objectives of the evaluation were to examine the effect of the first year of implementing the Mathematics Mastery programme on Year 7 pupils' maths test scores.

The process evaluation was conducted alongside quantitative evaluation to identify which elements of the Mathematics Mastery programme seemed important in supporting success.

Project team and roles

Helen Austerberry: Assisted with the testing in schools, ensuring high response rates.

Cosette Crisan: Assisted in the design of process evaluation instruments, conducted process evaluation data collection in two focus schools, and assisted in sense-checking findings from process evaluation.

Anne Ingold: Assisted with the testing in schools, ensuring high response rates.

John Jerrim (Principal Investigator): Led the trial design and data analysis, and wrote the final report. Overall management of the project.

Candia Morgan: Assistant lead in the process evaluation: assisted in the design of process evaluation instruments, conducted process evaluation data collection in one focus school, assisted in process evaluation analysis, and assisted in writing the process evaluation report.

Dave Pratt: Lead in the process evaluation: led in the design of process evaluation instruments, conducted all telephone interviews, led the process evaluation analysis, and led in writing the process evaluation report.

Cathy Smith: Assisted in the design of process evaluation instruments, conducted process evaluation data collection in two focus schools, and assisted in sense-checking findings from process evaluation.

Meg Wiggins: Assisted with the testing in schools, ensuring high response rates.

Ethical review

The evaluation was submitted to the Institute of Education ethics committee. Ethical approval was granted on 4th March 2013 (code FPS 462).

Impact evaluation methodology

Trial design

A clustered randomised controlled trial (RCT) has been used, with random allocation at the school level. The Mathematics Mastery programme involves a change to the maths curriculum within schools – therefore randomisation at either the pupil or class level was inappropriate. A total of 50 schools were recruited to participate in the trial during the September 2013 to August 2014 academic year. Half (25) of these schools were randomly allocated to intervention, with the remaining 25 schools allocated to control. All Year 7 pupils within the intervention schools received the Mathematics Mastery programme during the 2013/14 academic year. The control schools, on the other hand, were asked to proceed with 'business as usual'.

Eligibility

The charity delivering the intervention (Ark) was responsible for recruiting schools to participate in the trial. The main exclusion criteria were that schools could not already be receiving the Mathematics Mastery programme. Private (independent) schools were also ineligible for the study. Otherwise, Ark was free to recruit any secondary school within England. Schools were therefore deliberately selected – they cannot be considered a randomly chosen sample from a well-defined population. The majority of schools participating in the trial were from London or the South East (a significant minority were based in the Midlands). Ark was responsible for obtaining school-level consent to conduct the post-test within schools, along with the consent required to link the data collected to information from the National Pupil Database.

Intervention

Mathematics Mastery is a not-for-profit organisation working with primary and secondary schools across England. It was initially developed by Ark Schools to meet its aspiration to achieve success for every pupil.

With the UK currently placed 26th in PISA's international league table of maths attainment (OECD, 2013), the Mathematics Mastery team feel that more can be done to improve the quality of teaching in the UK. In particular, Ark believes there are four key issues facing UK schools: too many children falling behind, not enough children excelling, a focus on procedures over understanding, and negative attitudes towards maths as a subject.

The approach draws on a range of evidence-based practice from the UK and abroad. A key element of the approach is the introduction of a 'mastery curriculum', similar to those seen in cities that lead the international tables, such as Shanghai and Singapore (see Guskey, 2010). A 'mastery curriculum' is one in which the large majority of pupils progress through the curriculum content at the same pace, with differentiation through depth of exploration rather than accelerated content coverage. This approach reduces the need to repeatedly revisit topics, and promotes depth of understanding over memorised procedures. Consequently, fewer topics are covered in more depth within any given school year.

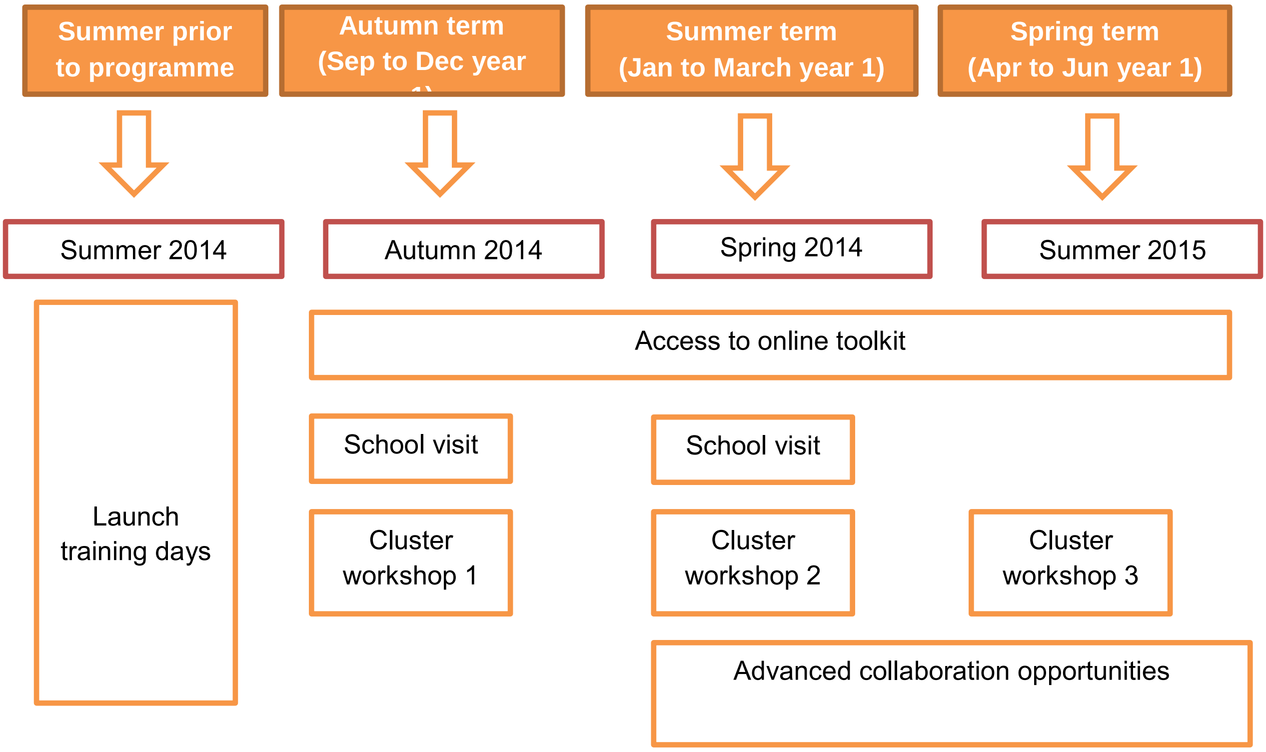
Adopting a mastery curriculum is challenging. To help teachers and schools make this shift, the programme provides a range of support:

- **Training and in-school support:** headteachers, maths coordinators and class teachers receive 1–2 days of launch training, two in-school development visits, three multi-school cluster workshops, and access to an online toolkit with continuous professional development (CPD) resources, assessments, and leadership frameworks.

- **Curriculum-embedded CPD:** teachers are supported to put the principles into day-to-day practice through high-quality mastery-aligned 'lesson designs', which they adapt to the needs of their class from the online toolkit.
- **Collaboration and peer support:** the programme brings together teachers in collaboration from different schools to develop practice (see Mulford, Silins, and Leithwood, 2004). This collaboration is both face to face and online, and focuses on sharing ideas and supporting each other in applying the approach. As all member schools are committed to the same curriculum framework, the relevance of other teachers' best practice is much more immediate. Teachers also work collaboratively with Mathematics Mastery to improve the programme itself year on year.

The diagram below gives an overview of the current support over the year-long programme:

Figure 1: Overview of the Mathematics Mastery Programme



In addition to the 'mastery curriculum', other features of the approach include a systematic approach to mathematical language (see Hoyles, 1985; Lee, 1998), frequent use of objects and pictures to represent mathematical concepts (see Heddens, 1986; Sowell, 1989), and an emphasis on high expectations (see Dweck, 2006; Boaler, 2010). The approach also aims to build fluency and understanding of underlying mathematical concepts in tandem, so that pupils gain deep conceptual understanding (see Skemp, 1976; Freudenthal, 1968). An example problem combining some of these elements can be found at the end of this summary. The approach takes a long-term view of transforming maths achievement. The curriculum is cumulative, to allow sufficient time for every child to access age-appropriate concepts and skills.

Since featuring in this evaluation, Mathematics Mastery has been accredited by the National Centre for Excellence in the Teaching of Mathematics (NCETM) for the quality of its CPD and has jointly won the TES Award for Maths and Numeracy (2014) with a partner school. According to Ark (the organisation delivering the intervention), last year 95% of headteachers and school leads using the Mathematics Mastery programme reported that the approach improved pupil attainment and the quality of teaching, and 98% said that it had increased pupils' enjoyment of mathematics.

The control condition was 'business as usual' within schools during the intervention year (September 2013 to August 2014). Control schools were then eligible to access the MM programme after the trial had ended (i.e. from September 2014 onwards).

The process evaluation found that the intervention approach had been largely adhered to by the schools, notwithstanding differing interpretations and levels of difficulty in adopting the project's intended teaching methods. Schools generally began by following the methods and materials as a detailed prescription. As their confidence with the materials increased, schools began to make choices, sometimes using alternative sources. Sometimes, schools began to identify difficulties in implementing the approach within the local circumstances and constraints, and so adapted the materials as the year progressed. Although schools mostly followed the approach advocated by the Mathematics Mastery programme, there were some changes which could be described as challenges to fidelity, and these are detailed later. For more detail, see the process evaluation section.

Example problem

The programme aims to build fluency and understanding of underlying mathematical concepts through practice and consolidation. For example, the Mathematics Mastery programme of study for Year 7 includes 25 hours of work on fractions, including multiplication and division of fractions.

One task that is offered during this topic is shown here. Pupils are asked to find which of the following expressions are equal to one-quarter.

a) 1/3 × 3/4	b) 1/8 × 1/2	c) 13/22 × 11/26	d) 6/7 × 11/26
e) 4/3 × 3/16	f) 15/28 × 7/15	g) 8/9 × 9/32	h) 9/10 × 5/18
i) 1/2 × 1/2	j) 5/3 × 3/20	k) 9/8 × 2/3	l) 1/4× 1
m) 1/2× 1/4 × 2	n) 1/2 × 1/2 ×4	o) 1/3 × 1/8 × 18/3	p) 1/3 × 1/1

This task is designed to offer pupils the practice they need to develop fluency in multiplying fractions, while also encouraging them to engage conceptually with both the relative size of fractions and the effect of multiplication.

For example:

Why is expression (a) equal to one-quarter? What models or diagrams might you use to convince someone of this?

Expression (e) is also equal to one-quarter. Compare (a) and (e). What is the same and what is different? If we already know that (a) is equal to one-quarter, how could we quickly know that (e) is also equal to one-quarter?

How do you know that expression (l) is equal to one-quarter? How can you use the formal method for multiplying fractions to calculate the product?

How can you know that expression (p) is not equal to one-quarter without calculating?

Higher-attaining pupils, who have completed the full set of practice exercises and engaged with the above considerations, would not be 'extended' by moving on to division of fractions, or to a new topic. Instead, they are asked when the following are equal to one-quarter, and why.

q) a/2 × 1/2a	r) 1/4 × b/c	s) d/4 × e/d

This requires pupils to generalise from their conceptual understanding of the fraction products.

Outcomes

The GL Assessment 'Progress in Maths' (PiM) 12 test (www.gl-assessment.co.uk/products/progress-maths) was used to examine children's skills in mathematics. This was administered using paper-and-pencil tests by class teachers within a standard maths lesson and took approximately one hour to complete. One teacher within each school was assigned as the lead contact, to whom the tests were sent. They were then responsible for ensuring the tests were completed by all Year 7 classes within the school, and for returning completed manuscripts to GL Assessments for marking. All scripts were marked by this independent organisation, which was blind to intervention. Schools were asked to complete this test during one week towards the end of the academic year (Monday 30th June–Friday 4th July 2014). However, due to timetable constraints, some flexibility was allowed around this date. For instance, one school in the control group requested that they complete the test three weeks earlier, while a number of other schools completed the tests a week later. The Progress in Maths test was chosen as it has been shown to provide a reliable measure of children's maths skills (see http://educationendowmentfoundation.org.uk/library/test-database for further details). Moreover, this test is *not* specific to the Mathematics Mastery intervention.

There are, however, a number of features of the Progress in Maths tests that have implications for the analysis and interpretation of results. First, certain questions within the Progress in Maths test cover material that is not contained within the Year 7 Mathematics Mastery curriculum. Specifically, around 40% of PiM test questions cover material that was not part of the Year 7 Mathematics Mastery curriculum. One would therefore expect children in the intervention group to be at a disadvantage relative to the control group on these particular questions (and within the calculator section of the Progress in Maths test as a whole). Second, the Progress in Maths test included both a calculator and a non-calculator section. However, as part of the Mathematics Mastery curriculum, Year 7 pupils do not work with calculators. Nevertheless, the evaluation team deemed that *overall* marks obtained across the *whole* Progress in Maths test (i.e. both calculator and non-calculator sections) would be the primary outcome. This is to ensure the evaluation is consistent with best practice—skills examined in the post-test should not be specific (or overly weighted towards) the intervention. However, it was agreed that a breakdown would also be presented by performance on (i) specific questions that were taught as part of the Mathematics Mastery curriculum[3] and (ii) the calculator versus non-calculator sections of the Progress in Maths test. Thus, in summary:

Primary outcome

- Total raw scores obtained across the whole of the Progress in Maths test

Secondary outcomes

- Total raw scores on the subset of questions covered in the Mathematics Mastery curriculum[4]
- Total raw scores on the calculator section of the Progress in Maths test
- Total raw scores on the non-calculator section of the Progress in Maths test

[3] These questions were chosen by Ark blind to the analysis.

[4] The 'fewer topics, greater depth' curriculum approach means Mathematics Mastery does not cover a range of topics on the test in Year 7 (and instead covers them in Years 8 and 9). The topics covered include symmetry, functions and graphs, negative numbers, coordinates, and 3D shapes. This secondary outcome looks at the remaining 60% of questions on the test that were covered as part of the Mathematics Mastery Year 7 curriculum.

Baseline test

Children's Key Stage 1 (KS1) and Key Stage 2 (KS2) test scores have been used to measure children's academic achievement prior to the Mathematics Mastery intervention. The former are based upon teacher assessments of pupils when they were aged 7. The latter are high-stakes,[5] externally marked tests conducted at the end of primary school when children are aged 11 (and took place in June 2013; three months before the intervention began). These baseline tests are used to (i) investigate balance between intervention and control groups in terms of prior attainment, and (ii) increase power and reduce possible confounding in the statistical analysis.

Sample size

Power calculations were based upon the following assumptions:

i. A school-level inter-cluster correlation (ICC) of 0.15 (i.e. $\rho = 0.15$).
ii. Equal cluster sizes of 200 Year 7 pupils per school.
iii. 50 schools (clusters) participating in the trial.
iv. 50% of the variance in the post-test is explained by the baseline covariates.
v. 80% power for a 95% confidence interval.

Based upon these assumptions, the trial would be able to detect an effect of 0.21 standard deviations.

Table 1 provides estimates of the ICC for the actual sample of schools/pupils that took part in the study. Estimates are presented for baseline (KS1 average points score and KS2 maths scores) and follow-up (PiM) tests, when using either a fixed or random school-level effect. Using a school-level fixed effect, the ICC was estimated to be 0.075 for KS1 average point scores, 0.056 for KS2 total maths scores, and 0.077 for the PiM tests. The analogous figures using a school-level random effect were 0.067 (KS1), 0.048 (KS2), and 0.069 (PiM).

Using actual data from the trial, points (i) to (v) above can be updated as follows:

i. A school-level ICC of approximately 0.07.
ii. Unequal school sizes, ranging from 46 to 234 pupils (median school = 126).
iii. After attrition of 6 schools, 44 schools participated in the trial.
iv. Two-thirds of the variance in the post-test is explained by the baseline covariates.
v. 80% power for a 95% confidence interval.

Using these figures, the actual minimum detectable effect was approximately 0.14.

[5] These tests are high-stakes, for schools at least, with information on performance made public and used to rank schools.

Table 1: Estimated inter-cluster correlation

	Fixed effect	Random effect
Key Stage 1 APS	**0.075**	**0.067**
Key Stage 2 Maths	**0.056**	**0.048**
PiM test scores	**0.077**	**0.069**

Notes: *Figures refer to the proportion of the variation in pupils' test scores occurring between schools. Estimates using KS1 APS (average point score) and KS2 maths total marks for pupils within all 50 schools were initially randomised. Estimates for PiM were based upon the 5,938 children within the 44 responding schools with data available.*

Randomisation

All aspects of the randomisation were conducted by the independent evaluation team. Schools were ranked by the percentage of children who achieved A*–C in GCSE English and Maths, using the most recent data available at the time of randomisation (this was the 2011/2012 academic year). Schools were then divided into pairs, and a random number generated for each using Excel. The school with the higher random number within each pair was assigned to intervention, and the school with the lower number within each pair to control.

Analysis

The impact of the intervention will be determined by the following regression model:

$$Y_{ij}^{Post} = \alpha + \beta.Treat_j + \gamma.Y_{ij}^{Pre} + \delta.C_{ij} + \varepsilon_{ij} \qquad (1)$$

Where:

Y^{post} = child's post-test score on the Progress in Maths test

Y^{pre} = child's baseline scores on the Key Stage 2 and Key Stage 1 tests

Treat = a binary variable indicating whether the child was enrolled in an intervention or control school (0 = control; 1 = intervention)

C = a series of additional control variables potentially associated with the outcome (e.g. gender, Free School Meals, ethnicity)[6]

ε = error term (with children clustered within school)

i = child i

j = school j

To account for the clustering of pupils within schools, the STATA survey (svy) command was used to make Huber–White adjustments to the estimated standard errors. The coefficient of interest from equation (1) is β—is there a positive effect of the MM intervention?

After our main analysis, we re-estimated model 1: (a) restricting the sample to FSM children only, and (b) using quantile regression to investigate differences in intervention effects across the achievement distribution.

[6] The study protocol did not explicitly mention controlling for factors other than prior achievement. However, it was decided that this was the most appropriate approach in order to (i) maximise statistical power and (ii) account for any chance differences occurring between intervention and control groups in terms of observable characteristics.

Standardisation of PiM maths test scores (Hedges' g)

Post-test scores are converted into z-scores, using the following formula:

$$Z_{ij} = \frac{(X_{ij} - \bar{X})}{SD_{Pool}}$$

Where:

X = children's post-test score in the analysis sample

$\bar{X}$ = the mean post-test score of all children in the analysis sample

Z = the standardised post-test score

SD_{Pool} = the standard deviation of post-test scores (pooled across all children in the analysis sample)

i = pupil i

j = school j.

Following EEF guidelines, the pooled standard deviation has been calculated using Hedges' G:

$$G = \sqrt{\left[\frac{(n_T - 1).S_T^2 + (n_C - 1).S_C^2}{n_T + n_C - 2}\right]} * \left[1 - \frac{3}{4(n_T + n_c) - 9}\right]$$

Where:

G = Hedges' G

n_T = number of observations in the intervention group

S_T = standard deviation of post-test scores for the intervention group

n_C = number of observations in the control group

S_C = standard deviation of post-test scores for the control group.

The use of Hedges' G may be a somewhat conservative choice. Nevertheless, an alternative standard deviation (the simple pooled standard deviation across all children with a post-test score) has also been used. There was very little change in results.

Missing covariate data

The robustness of results is considered using different ways of handling missing covariate information. This includes (i) a complete case analysis, (ii) 'missing' dummy variables, and (iii) multiple imputation by chained equations (MICE).

Protocol, registration, and data availability

The protocol for this study is published online at:

http://educationendowmentfoundation.org.uk/projects/maths-mastery-secondary/

The trial has been registered with the independent ISRCTN website at:

http://controlled-trials.com/ISRCTN70922140/

The trial registration number is ISRCTN70922140 and the DOI is 10.1186/ISRCTN70922140.

Data from the trial will be submitted to the EEF database upon final publication of the report. The data will be available for long-term follow-up of participants through to GCSE.

Process evaluation methodology

Methods

The Mathematics Mastery documentation was analysed in order to identify key issues which the intervention stressed and was intended to have an impact upon. Using their teaching experience and knowledge of the research literature, the process evaluation team identified related issues and developed a bespoke baseline questionnaire. The Mathematics Mastery team was invited to respond to the design of the questionnaire, which was further developed in the light of their comments. With the help of the intervention team, all teachers of Year 7 mathematics in all the intervention and control schools were requested to complete the questionnaire prior to the start of the intervention. The questionnaire was made available online.

A very similar online final questionnaire was designed to capture data around the same issues towards the end of the intervention.

The main effort of the process evaluation was directed towards five focus schools chosen from the intervention group. (See below for an explanation of how these schools were chosen.)

Focus group discussions with teachers from the five focus schools involved in implementing the intervention were held prior to the start of the intervention and towards the end of the intervention in each focus school. A schedule was created by the process evaluation team to identify key questions and possible follow-up questions around which the focus group discussion was based. These questions were based on the same key issues identified above.

It was planned that two classroom lessons were observed in each focus school prior to the intervention and a further two lessons were observed towards the end of the intervention. (In practice, logistics prevented us from observing classes at one school in the final set of observations, and it proved convenient to observe a third classroom in another school.) An observation template was designed to structure the comments made by the observers around the same key issues.

Telephone interviews of 30–40 minutes were conducted with the leader of the Mathematics Mastery teaching and with one other Year 7 teacher who was also involved in the programme of teaching for each of the five focus schools.

Choice of focus schools

The baseline questionnaire responses and other profile data were used to select the five focus schools. First, we only considered schools where at least six teachers had made responses to the questionnaire. This was a pragmatic criterion based on the need to ensure a suitably large focus group, flexibility in choice of teachers to observe, and further options should teachers leave during the intervention year.

We then looked for candidate schools on the basis of variety across a range of criteria: (i) large schools and relatively small schools (given that we only considered schools with at least six responses); (ii) schools with high and schools with low proportions of free school meals; (ii) schools with poor or schools with good progress in mathematics according to their most recent Ofsted reports. We then chose from the candidate schools according to logistical convenience for the team members, who would be visiting the schools on a number of occasions. The five schools selected fulfilled these criteria as below:

School reference	Size	FSM	PiM
A	Large	High	Good
B	Large	Low	Poor
C	Small	High	Poor
D	Small	Medium	Poor
E	Large	Medium	Good

Data collection

The questionnaires were made available in a timely way. For the baseline questionnaire, responses were received from 214 teachers from across 47 schools (94% of all control and intervention schools). Of these, 110 were from control schools and 104 were from intervention schools, though the teachers did not know at the time of completion whether they were in an intervention or control school. In the final questionnaire, responses were received from 138 teachers from across 45 schools (90%). Of these, 66 were from the control schools and 72 were from intervention schools. The reduced number of teachers responding to the final questionnaire was predictable. Considerable efforts were made with very welcome support from the Mathematics Mastery project team to maximise the response rate. In view of the inevitable focus on other matters at the end of a school year and the fact that the control group schools had not been directly involved in the intervention, we were pleased and a little surprised that such a good response had been generated.

Each focus group was led by one of the members of the process evaluation team. The size of the focus group varied but was typically about six people. The discussion was recorded and transcribed by a professional transcriber. There was some loss of data when comments could not be heard sufficiently clearly, but this was very limited and it was felt that a faithful record of each discussion had been captured.

Using the specially designed templates, observers from the process evaluation team recorded field notes of the activity during the classroom observations. Each observer wrote a summary immediately after the lesson using the key issues that had been previously identified. Photographs were taken of relevant materials.

The telephone interviews were recorded and transcribed by a professional transcriber. The quality of the recordings was very good and there was almost no loss of data.

Analysis

For the baseline focus group discussions and classroom observations, both analysts independently analysed the data from school A in order to establish common interpretations of the responses to the questions asked and the key issues identified. The remaining schools were then shared between them. The first stage was to summarise the responses to each question. The second stage was to analyse these summaries to identify evidence relating to each of the key issues. The third stage was to identify evidence related to the key issues in the observation field notes and supporting data. This process was repeated for the final focus group discussions and lesson observations. Finally, the key issues were shared between the two analysts who looked across the schools to identify similarities and differences with respect to each key issue.

The telephone interview transcripts were read in the light of the above analysis. Where there were discrepancies between the telephone interview recordings and the findings emerging from the analysis of the focus group discussions and observations, these were noted. Where further evidence in support of the findings was found, this was also noted. The two analysts then discussed the list of discrepancies and further evidence. Both were used as appropriate to modify the findings. Where new issues were thrown up by the telephone interviews, the focus group discussions and observation data were re-inspected to decide how to address that issue. In practice, this was a rare occurrence, with broad agreement between the telephone interviews and the findings that had previously emerged being apparent.

With respect to the questionnaire data, the most powerful type of analysis would be to look for changes in individual teachers who completed the questionnaire before and after the intervention. Because of staff turnaround and in some cases non-completion of the questionnaires, there were only 27 teachers from the intervention schools who completed both questionnaires and so could be analysed in this way. The findings are therefore very tentative but may nevertheless be of interest in triangulating findings from other sources. We have reported such findings where relevant, whether in support of or in contradiction to the conclusions drawn from other sources.

Because the number of teachers who completed both questionnaires was small, we also analysed the whole set of intervention and control teachers to look for comparative trends in differences between the extent to which practice seemed to have changed across the period of the intervention. Of course, many of the teachers were not the same people who completed the questionnaire, so the inferences are again tentative, though some confidence can be drawn when the conclusions for the 27 teachers who completed both questionnaires are supported by similar findings from the analysis of the intervention and control groups as a whole. Any such findings are reported where relevant, whether in support of or in contradiction to the conclusions drawn from other sources.

Impact evaluation results

Timeline

Schools recruited = February–April 2013

Schools assigned to intervention or control group = 2nd May 2013

Key Stage 2 tests conducted = June 2013

Mathematics Mastery programme delivered in intervention schools between September 2013 and July 2014

PiM post-test week: Monday 30th June–Friday 4th July 2014

Participants

Sample allocation

Figure 2 provides details of sample allocation and attrition. A total of 50 schools were recruited to participate in the trial. Schools were evenly split between intervention (n = 25) and control (n = 25) groups.

All Year 7 children enrolled in one of the 50 participating schools in the trial on 3rd October 2013 were considered to be part of the Mathematics Mastery trial. (This was the date of the autumn school census in 2013.) Information on school enrolment on this date has been drawn directly from the National Pupil Database (NPD). A total of 4,004 children were enrolled in the 25 intervention schools and 3,708 in the 25 control schools.

Adherence

Three out of the 25 schools assigned to intervention dropped out of the Mathematics Mastery programme (although one of these schools still completed the post-test). However, as we are undertaking an Intention-To-Treat (ITT) analysis, these schools will continue to be included in the analysis, within their initially allocated group. There was no evidence that the MM programme was introduced in any control school during the trial period.

Missing data at baseline

Pupils' KS1 and KS2 test scores were drawn from the National Pupil Database (NPD). Information was missing for a small number of pupils not enrolled in a school in England when these tests were conducted, or where there were problems linking NPD data over time. Overall, Key Stage 1 average point scores were available for 6,900 (89%) of the 7,712 children participating in the trial. Key Stage 2 total maths scores were available for 7,254 pupils (94%).

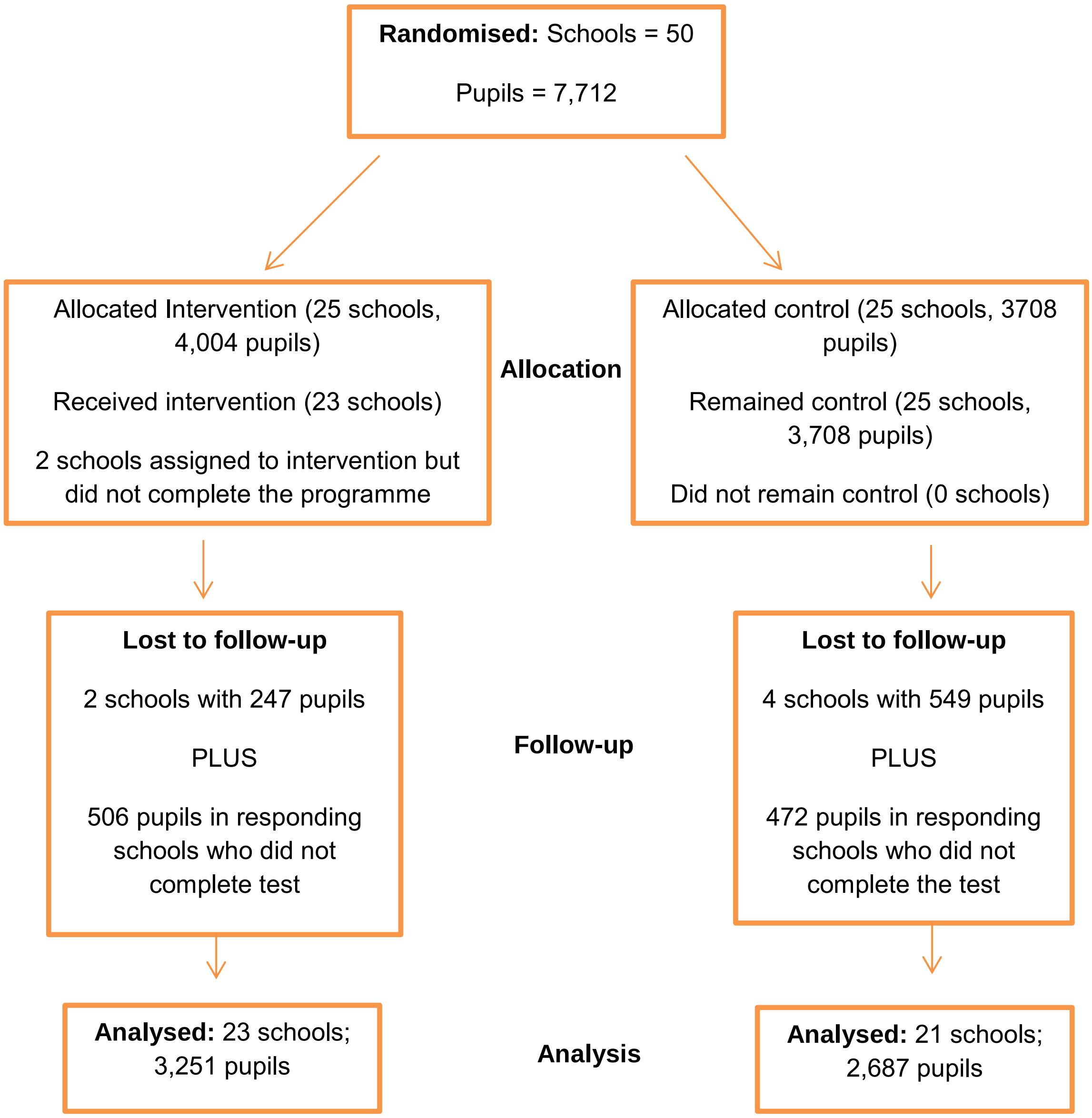
Attrition

Six schools (two intervention schools and four control schools) did not complete the post-test. Hence the cluster (school) level response rate was 88% (84% for the control group and 92% for the intervention group). This reduced the number of pupils to 3,757 in the intervention group and 3,159 in the control group.

Moreover, within the 44 schools that did conduct the post-test, test scores were missing for some individual pupils. The most likely reasons are that these children were absent on the day the school conducted the post-test (e.g. through illness) or that they had left the school completely (e.g. moved to another school). The pupil response rate (within the 44 schools who conducted the post-test) was

85% in the control group and 87% in the intervention group. After excluding children with missing post-test data, pupil sample sizes fell to 3,251 in the intervention group (81% of the original allocation) and 2,687 in the control group (72% of the original allocation). These 5,938 children form the analysis sample.

Figure 2: Sample allocation and attrition



Non-response analysis

The NPD can be used to compare the characteristics of respondents and non-respondents across the intervention and control groups. Results are presented in Tables 2 and 3. The former illustrates that children who did not complete the post-test tend to have lower levels of prior achievement. This was particularly true for pupils within the intervention group. For instance, non-respondents from the intervention group scored (on average) 0.24 standard deviations below the sample mean on the Key Stage 2 maths test. This compares to 0.04 standard deviations above the mean for respondents in the intervention group. Analogous figures for the control group were -0.01 and 0.02 standard deviations respectively. Similar findings hold for other pre-test scores, including Key Stage 2 reading scores and Key Stage 1 average points scores. Moreover, Table 3 suggests that boys and children in receipt of FSM were also more likely to have missing post-test data than their female, non-FSM counterparts. Specifically, 37% of intervention group non-respondents were eligible for FSM, compared to just 28% of respondents. Likewise, 52% of control group respondents were male, compared to 58% of non-respondents.

Together, Tables 2 and 3 suggest that attrition from the sample is not random. Rather, lower-achieving, disadvantaged boys were more likely to have dropped out of the study than other groups. Moreover, there is some evidence that the nature of the attrition was slightly more selective in the intervention arm. It will therefore be important to compare the balance of observable characteristics between intervention and control groups both before and after attrition has been taken into account.

Table 2: A comparison of prior achievement between children who did and who did not complete the post-test

	Intervention		Control	
	Respondent	Non-respondent	Respondent	Non-respondent
Key Stage 1 maths				
Level 1 %	11	18	11	14
Level 2A %	28	20	26	23
Level 2B %	27	27	29	28
Level 2C %	19	26	19	20
Level 3 %	15	9	16	14
Key Stage 1 reading				
Level 1 %	16	26	16	21
Level 2A %	24	17	23	22
Level 2B %	25	25	28	25
Level 2C %	16	18	15	14
Level 3 %	19	14	19	18
Key Stage 1 writing				
Level 1 %	20	29	20	24
Level 2A %	19	14	16	18
Level 2B %	27	22	31	26
Level 2C %	25	30	26	25
Level 3 %	9	5	7	8
KS1 APS (standardised)	0.065	–0.339	0.041	–0.087
KS2 maths score (standardised)	0.036	–0.244	0.022	–0.009
KS2 reading score (standardised)	0.045	–0.207	0.025	–0.072
Pupil n	**3,251**	**753**	**2,687**	**1,021**

Notes: *Figures reported for children with complete Key Stage 1 or Key Stage 2 data.*

Table 3: A comparison of demographic characteristics between children who did and who did not complete the post-test

	Intervention		Control	
	Respondent	**Non-respondent**	**Respondent**	**Non-respondent**
Eligible for FSM				
No %	72	63	74	68
Yes %	28	37	26	32
Gender				
Female %	51	39	48	42
Male %	49	61	52	58
Ethnic group				
White %	47	49	51	47
Asian %	26	23	13	13
Black %	17	16	21	21
Mixed %	6	8	7	11
Chinese %	0.4	0.1	0	1
Other/unclassified %	4	4	7	7
Pupil n	**3,251**	**753**	**2,687**	**1,021**

Notes: *Figures reported for children with complete Key Stage 1 or Key Stage 2 data.*

Pupil characteristics

As randomised

First, we investigate whether balance was achieved among all 7,712 pupils within the 50 schools as they were initially randomised (i.e. possible imbalance due to attrition and non-response is, for the moment, not taken into account). Table 4 compares KS1 scores for children in the intervention and control groups across three subject areas (numeracy, reading, and writing) and average point scores (standardised to mean 0 and standard deviation 1). The difference in KS1 APS between intervention and control groups is small, standing at just 0.01 standard deviations. Similarly, the distribution of KS1 maths scores is very similar across the two groups, with differences at any given level typically just one or two percentage points. Similar findings also hold for children's KS2 maths and English test scores. In particular, average Key Stage 2 maths test scores differ between the intervention and control groups by just 0.03 standard deviations (among the 7,254 children where this information is available). Likewise, the difference in KS2 English test scores is less than 0.01 standard deviations. Overall, Table 4 suggests that the intervention and control groups are well balanced in terms of prior academic achievement.

Table 5 considers balance between intervention and control groups in terms of other observable characteristics. (These characteristics are presented for all 7,712 children initially randomised.) There are more Asian children (26% versus 13%) and children eligible for FSM (30% versus 28%) in the intervention group than the control group. There also appears to be slightly more males than females in the control group (46% female) compared to the intervention group (49% female). Nevertheless, most of the differences observed in Table 5 are relatively small, suggesting that there is also reasonable balance on a range of other baseline characteristics.

Table 4: A comparison of baseline (KS1 and KS2) test scores between intervention and control groups *as initially randomised*

	Control	Intervention
Key Stage 1 maths		
Level 1 %	10	11
Level 2A %	22	24
Level 2B %	25	24
Level 2C %	17	18
Level 3 %	13	12
Missing %	13	12
Key Stage 1 reading		
Level 1 %	15	16
Level 2A %	20	20
Level 2B %	23	22
Level 2C %	13	14
Level 3 %	16	16
Missing %	14	13
Key Stage 1 writing		
Level 1 %	18	19
Level 2A %	14	16
Level 2B %	25	23
Level 2C %	22	22
Level 3 %	6	7
Missing %	15	14
KS1 APS (standardised)	0.007	–0.006
KS2 maths score (standardised)	0.014	–0.013
KS2 reading score (standardised)	–0.001	0.001
School n	**25**	**25**
Pupil n	**3,708**	**4,004**

Notes: *Authors' calculations using the NPD. KS1 APS and KS2 scores in the lower panel have been standardised to have a mean of 0 and a standard deviation of 1 (across pupils within the 50 schools initially randomised). Figures reported for children where data available.*

Table 5: A comparison of children in the intervention and control groups in terms of observable characteristics as initially randomised

	Control	Intervention
Eligible for FSM		
No %	72	70
Yes %	28	30
Gender		
Female %	46	49
Male %	54	51
Ethnic group		
White %	50	47
Asian %	13	26
Black %	21	16
Mixed %	8	7
Chinese %	0.3	0.4
Other/unclassified %	6.8	3.9
School n	25	25
Pupil n	3,708	4,004

Notes: *Authors' calculations using the NPD. Figures reported for children where data available.*

Analysis sample

We now re-present our comparison of the intervention and control groups, having restricted our attention to the analysis sample of 5,938 pupils. Results are presented in Tables 6 and 7. Despite certain groups being more likely to drop out of the study than others (recall Tables 2 and 3), there remains a good balance between intervention and control. For instance, standardised Key Stage 2 maths test scores equal 0.022 for the control group and 0.036 for the intervention group. There is thus a minimal difference of just 0.014 standard deviations. Similarly, there is a difference of just 0.02 standard deviations in terms of Key Stage 2 reading test scores and Key Stage 1 average point scores. Moreover, Table 7 continues to suggest similar proportions of children eligible for FSM (26% versus 28%) and girls (48% versus 51%) within the two groups. Thus the only notable difference remains the greater number of Asian children in intervention schools (26%) compared to control schools (13%). (This seems to have occurred by chance in the randomisation, not due to non-random attrition from the study). Nevertheless, Tables 6 and 7 generally suggest that individual observable characteristics are well balanced between the intervention and control groups, even after attrition from the sample has been taken into account.[7] (Appendix A provides an investigation of balance at the school (cluster) level.)

7 We note that although balance seems good for any given characteristic, taken together there may be some evidence of a slight trend across characteristics of a small advantage of the intervention group.

Table 6: A comparison of baseline (KS1 and KS2) test scores between intervention and control groups (analysis sample)

	Control	Intervention
Key Stage 1 maths		
Level 1 %	10	10
Level 2A %	23	25
Level 2B %	25	24
Level 2C %	17	17
Level 3 %	14	13
Missing %	12	10
Key Stage 1 reading		
Level 1 %	14	15
Level 2A %	20	21
Level 2B %	24	22
Level 2C %	13	14
Level 3 %	16	17
Missing %	12	11
Key Stage 1 writing		
Level 1 %	17	18
Level 2A %	14	17
Level 2B %	27	24
Level 2C %	23	22
Level 3 %	6	8
Missing %	13	12
KS1 APS (standardised)	0.041	0.065
KS2 maths score (standardised)	0.022	0.036
KS2 reading score (standardised)	0.025	0.045
School n	21	23
Pupil n	2,687	3,251

Notes: *Authors' calculations using the NPD. KS1 APS and KS2 scores in the lower panel have been standardised to have a mean of 0 and a standard deviation of 1 (across pupils within the 50 schools initially randomised). Figures reported for children where data available.*

Table 7: A comparison of children in the intervention and control groups in terms of observable characteristics (analysis sample)

	Control	Intervention
Eligible for FSM		
No %	**74**	**72**
Yes %	**26**	**28**
Gender		
Female %	**48**	**51**
Male %	**52**	**49**
Ethnic group		
White %	**51**	**47**
Asian %	**13**	**26**
Black %	**21**	**17**
Mixed %	**7**	**6**
Chinese %	**0**	**0**
Other/unclassified %	**7**	**4**
School n	**21**	**23**
Pupil n	**2,687**	**3,251**

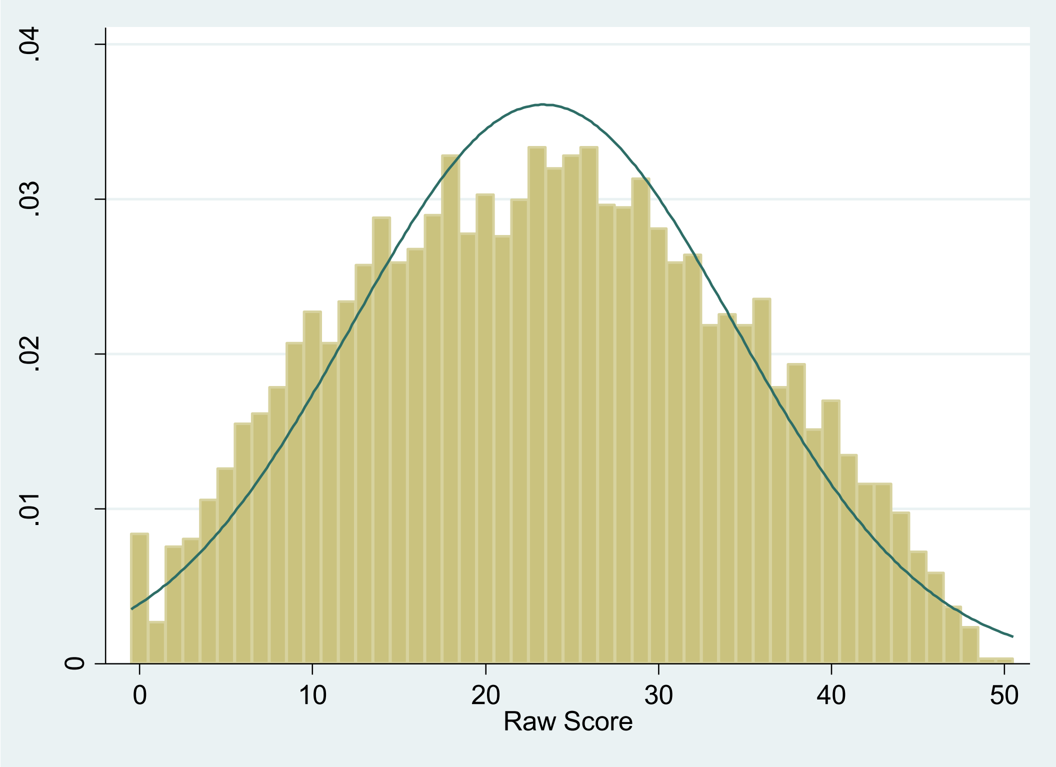
Notes: *Authors' calculations using the NPD. Figures reported for children where data available.*

Outcomes and analysis

Descriptive statistics

Figure 3 plots the distribution of PiM (post-test) scores for the 5,938 children in the analysis sample. There is little evidence of either floor or ceiling effects, with the distribution being broadly symmetric around the mean of 23 points (standard deviation 11 points). Figure 4 then plots the correlation between the main pre-test (KS2 maths) and the PiM post-test. Pre and post-test performance are clearly highly associated, with the estimated Pearson correlation coefficient standing at 0.77 (Spearman's rank = 0.79). However, Figure 4 also suggests that the relationship between the pre- and post-test is non-linear, with a quadratic perhaps being the more appropriate functional form. Finally, we estimate an Ordinary Least Squares (OLS) regression model with total (raw) PiM scores as the dependent variable and the baseline characteristics as covariates. Parameter estimates are presented in Table 8. The R^2 is 0.66, suggesting that two-thirds of the variance in the post-test can be explained by children's baseline characteristics.

Figure 3: The distribution of children's Progress in Maths (post-test) total raw scores



Notes: *Authors' calculations. Sample restricted to analysis sample of 5,938 observations. 'Density' refers to the probability density function of PiM post-test scores. It illustrates the proportion of children obtaining each raw score.*

Figure 4: Scatter plot between children's Key Stage 2 maths (pre-test) score and Progress in Maths (post-test) score

Notes: *Standardised Key Stage 2 maths test scores are plotted along the horizontal axis. Progress in Maths total raw score is plotted on the vertical axis. A non-parametric locally weighted regression line is plotted to illustrate the relationship between the two. The Pearson correlation coefficient = 0.77 and Spearman's rank = 0.79.*

Table 8: Regression estimates of post-test (PiM) score against the baseline covariates

	Beta	SE
Key Stage 1 maths (Ref: Level 1)		
Level 2a	3.83**	0.64
Level 2b	2.53**	0.51
Level 2c	1.22**	0.44
Level 3	5.64**	0.90
Key Stage 1 average points score	0.52*	0.27
Key Stage 2 maths test score	7.07**	0.20
Key Stage 2 maths test score squared	1.31**	0.11
Key Stage 2 English score	1.21**	0.15
FSM (Ref: No)		
Yes	-1.15**	0.23
Ethnic group (Ref: Group)		
Other ethnic group	1.16*	0.64
Asian	0.58	0.55
Black	–0.94**	0.42
Chinese	0.77	2.47
Mixed	0.63	0.46
Unclassified	1.37	0.87
Missing	2.63*	1.44
Gender (Ref: Female)		
Male	–1.33**	0.24
Constant	20.30**	0.67
R-Squared	0.666	
N	5,616	

Notes: *Authors' calculations. Sample size equals 5,616 rather than 5,938 due to missing data on selected covariates. Dependent variable is total raw PiM score (unstandardised). SE stands for standard error. * and ** indicate statistical significance at the 10% and 5% levels.*

Regression results: primary outcome

The primary dependent variable used in the OLS regression model is the standardised PiM test score (recall that this has been standardised in terms of Hedges' G). Parameter estimates are presented in Table 9. The left-hand columns present results for all children, while the right-hand columns restrict the analysis to FSM pupils only.[8] The highlighted row provides the impact of the Mathematics Mastery 'intervention' in terms of an effect size (Hedges' G).

Children who received the Mathematics Mastery intervention scored, on average, +0.055 standard deviations higher on the PiM post-test. This did not reach statistical significance at conventional thresholds ($t = 1.20$; $p = 0.24$), with the 95% confidence interval ranging from –0.037 to +0.147. Turning to the FSM-only sample, the estimated effect size is +0.066 with the 95% confidence interval

*8 In supplementary analysis, we have estimated regression models using the 'all children' sample including an FSM * Intervention interaction. Substantive conclusions remain largely unchanged.*

ranging from –0.037 to +0.169 ($p = 0.21$). Moreover, we also estimated a model including a FSM-by-intervention interaction. Results suggested there was little evidence of heterogeneous intervention effects by FSM. Consequently, although the Mathematics Mastery intervention may have had a small positive effect on overall PiM test scores, one cannot rule out the possibility that this finding is due to sampling variation.

Table 9: Primary outcome (total PiM score) regression results

	All children		FSM only	
	Beta	SE	Beta	SE
Intervention group (Ref: Control)				
Intervention	0.055	0.046	0.066	0.051
Key Stage 1 maths (Ref: Level 1)				
Level 2a	0.21**	0.05	0.36**	0.08
Level 2b	0.14**	0.04	0.22**	0.06
Level 2c	0.06*	0.03	0.07	0.05
Level 3	0.32**	0.06	0.48**	0.11
Key Stage 1 average points score	0.09**	0.02	0.08**	0.04
Key Stage 2 maths test score	0.66**	0.02	0.62**	0.03
Key Stage 2 maths test score squared	0.12**	0.01	0.12**	0.01
Key Stage 2 English score	0.11**	0.01	0.10**	0.02
FSM (Ref: No)				
Yes	–0.10**	0.02	-	-
Ethnic group (Ref: Other)				
Asian	–0.08	0.06	–0.07	0.08
Black	–0.19**	0.05	–0.17**	0.06
Chinese	0.00	0.21	0.07	0.15
Mixed	–0.03	0.06	–0.06	0.08
Unclassified	0.08	0.09	0.09	0.14
White	–0.08	0.05	–0.13**	0.06
Gender (Ref: Female)				
Male	–0.11**	0.02	–0.14**	0.04
English as additional language (Ref: No)				
Yes	0.05**	0.03	0.12**	0.04
Constant	–0.15**	0.06	–0.35**	0.08
N	5,919		1,610	

*Notes: Authors' calculations. Sample size equals 5,919 rather than 5,938 as covariates could not be imputed for 19 observations. Dependent variable is total PiM test score (standardised to mean 0 and standard deviation 1). SE stands for standard error. * and ** indicate statistical significance at the 10% and 5% levels. Standard errors clustered at the school level.*

Secondary outcomes

As part of the Mathematics Mastery programme, children complete a different Year 7 curriculum than the status quo. A key element is that they no longer work with calculators—something that is somewhat different to the 'standard practice' likely to be followed by the control group. This has important implications for how one interprets the primary outcome results presented in the sub-section above. Specifically, the PiM test includes both a calculator and a non-calculator section, with overall scores based approximately one-third on the former and two-thirds on the latter.[9] Yet there is little reason to believe that children in the Mathematics Mastery intervention group would perform better on the calculator section of the PiM test. Indeed, for the reasons given above, they may well do worse. Consequently, we now consider how children performed on the different sections of the PiM test.

Table 10 presents results from an OLS regression model where children's calculator score is the dependent variable. As with total scores, this has been standardised to mean 0 and standard deviation 1, with the latter based upon Hedges' G. Estimates presented in this table can thus be interpreted in terms of an effect size. Interestingly, the effect of the MM intervention is essentially zero, with the 95% confidence interval ranging from –0.07 to +0.08. Similar findings hold for FSM children, with the estimated effect size standing at +0.03 (95% confidence interval, ranging from –0.07 to +0.13). Thus, despite not working with calculators as part of the Year 7 curriculum, there is no evidence that children in receipt of the Mathematics Mastery intervention have weaker skills in this area.

[9] The total (raw score) marks available on the PiM test was 50: 18 marks were available in the calculator section and 32 marks in the non-calculator section.

Table 10: Regression results: children's calculator scores

	All children		FSM only	
	Beta	**SE**	**Beta**	**SE**
Intervention group (Ref: Control)				
Intervention	**0.004**	**0.038**	**0.030**	**0.048**
Key Stage 1 maths (Ref: Level 1)				
Level 2a	**0.17****	**0.05**	**0.32****	**0.08**
Level 2b	**0.14****	**0.04**	**0.25****	**0.06**
Level 2c	**0.06**	**0.03**	**0.09**	**0.06**
Level 3	**0.26****	**0.06**	**0.36****	**0.12**
Key Stage 1 average points score	**0.11****	**0.02**	**0.11****	**0.04**
Key Stage 2 maths test score	**0.65****	**0.02**	**0.59****	**0.04**
Key Stage 2 maths test score squared	**0.12****	**0.01**	**0.12****	**0.01**
Key Stage 2 English score	**0.08****	**0.01**	**0.08****	**0.03**
FSM (Ref: No)				
Yes	**–0.08****	**0.02**	**-**	**-**
Ethnic group (Ref: Other)				
Asian	**–0.08**	**0.06**	**–0.03**	**0.07**
Black	**–0.13****	**0.04**	**–0.05**	**0.06**
Chinese	**0.17**	**0.15**	**0.12**	**0.32**
Mixed	**–0.03**	**0.06**	**–0.02**	**0.09**
Unclassified	**0.09**	**0.12**	**0.10**	**0.17**
White	**–0.06**	**0.05**	**–0.08**	**0.06**
Gender (Ref: Female)				
Male	**–0.05****	**0.02**	**–0.06**	**0.04**
English as additional language (Ref: No)				
Yes	**0.09****	**0.02**	**0.10****	**0.05**
Constant	**–0.17****	**0.06**	**–0.38****	**0.10**
N	**5,887**		**1,598**	

*Notes: Authors' calculations. Sample size equals 5,887 rather than 5,938 as covariates could not be imputed for 19 observations and the calculator score was missing for 32 observations. The dependent variable is children's score in the calculator section on the PiM test (standardised to mean 0 and standard deviation of 1). All figures refer to effect sizes. SE stands for standard error. * and ** indicate statistical significance at the 10% and 5% levels. Standard errors are clustered at the school level.*

Table 11 turns to results for the remaining (non-calculator) component of the PiM test. Again, the dependent variable has been converted into a z-score. One now observes a slightly larger impact of the intervention (effect size = +0.08) which sits on the boundary of statistical significance at the 10% level (t = 1.56; p = 0.13; 95% confidence interval = –0.02 to +0.18). The effect is very similar for the FSM-only sample (effect size +0.08; 95% confidence interval = –0.03 to +0.18).

Recall that the Mathematics Mastery intervention is particularly concerned with the 'mastery' of, and raising the attainment of, low achievers. Thus one might anticipate the intervention to be particularly effective in the bottom half of the test score distribution. In exploratory analysis,[10] Figure 5 examines whether there is any evidence of heterogeneity in intervention effect across the non-calculator test score distribution. Specifically, this figure presents estimates from a series of quantile regression models, with estimates produced at every 5th percentile between p15 and p85. (A full set of parameter estimates can be found in Appendix B.)

The quantile regression estimates presented in Figure 5 suggest that the effect of the intervention is approximately +0.10 between p15 and p50. These estimates are almost all significantly greater than zero at the 5% (p15, p20, p25) or 10% (p30, p35, p40, p50) level. In contrast, the effect of the intervention is notably smaller in the top half of the non-calculator test score distribution (approximately +0.05 standard deviations). Indeed, above p55, none of the quantile regression estimates reaches statistical significance at the 10% level.

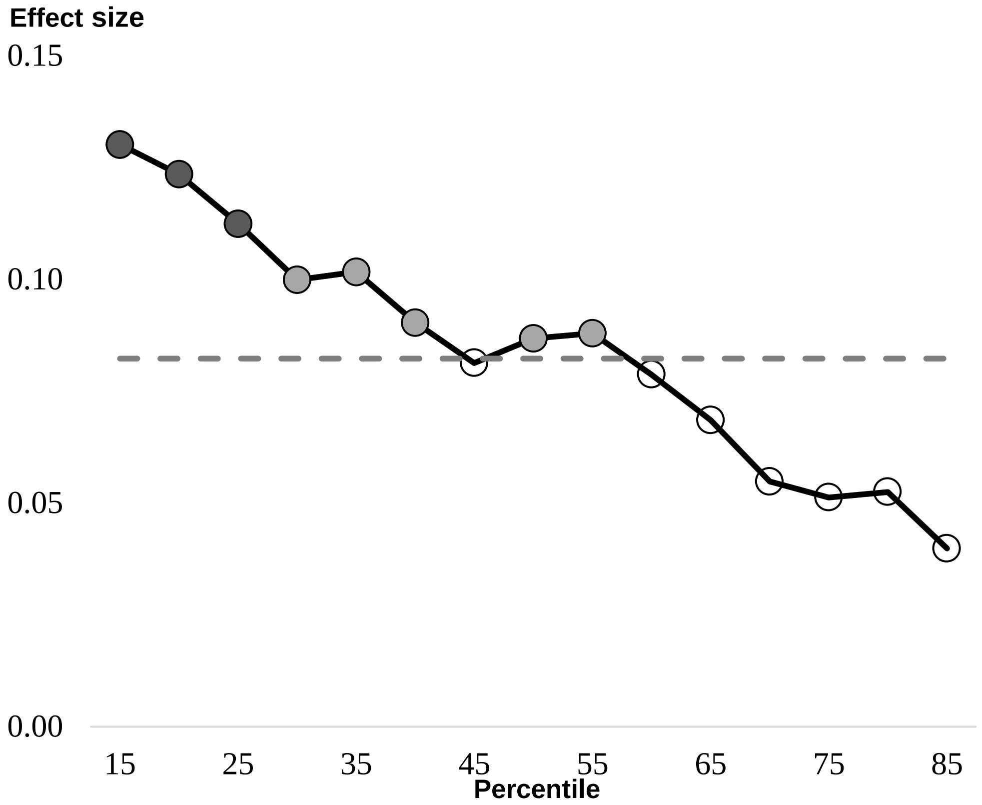
[10] This is 'exploratory' in the sense that the use of quantile regression was not explicitly stated in the study protocol.

Table 11: Regression results: children's non-calculator scores

	All children		FSM only	
	Beta	**SE**	**Beta**	**SE**
Intervention group (Ref: Control)				
Intervention	**0.077**	**0.049**	**0.075**	**0.053**
Key Stage 1 maths (Ref: Level 1)				
Level 2a	**0.22****	**0.05**	**0.41****	**0.08**
Level 2b	**0.13****	**0.04**	**0.23****	**0.07**
Level 2c	**0.05**	**0.03**	**0.08**	**0.06**
Level 3	**0.32****	**0.06**	**0.55****	**0.11**
Key Stage 1 average points score	**0.08****	**0.02**	**0.04**	**0.04**
Key Stage 2 maths test score	**0.64****	**0.02**	**0.59****	**0.04**
Key Stage 2 maths test score squared	**0.11****	**0.01**	**0.11****	**0.01**
Key Stage 2 English score	**0.12****	**0.02**	**0.11****	**0.02**
FSM (Ref: No)				
Yes	**–0.10****	**0.02**	**-**	**-**
Ethnic group (Ref: Other)				
Asian	**–0.08**	**0.06**	**–0.07**	**0.09**
Black	**–0.21****	**0.05**	**–0.21****	**0.07**
Chinese	**0.20****	**0.09**	**0.04**	**0.12**
Mixed	**–0.02**	**0.06**	**–0.06**	**0.08**
Unclassified	**0.07**	**0.08**	**0.06**	**0.13**
White	**–0.09***	**0.05**	**-0.15****	**0.07**
Gender (Ref: Female)				
Male	**–0.12****	**0.02**	**–0.16****	**0.04**
English as additional language (Ref: No)				
Yes	**0.01**	**0.03**	**0.11****	**0.04**
Constant	**–0.11***	**0.06**	**–0.33****	**0.09**
N	**5,871**		**1,591**	

*Notes: Authors' calculations. Sample size equals 5,871 rather than 5,938 as covariates could not be imputed for 19 observations and the non-calculator score was missing for 48 observations. The dependent variable is the children's score in the non-calculator section of the PiM test. (Standardised to mean 0 and standard deviation of 1.) All figures refer to effect sizes. SE stands for standard error. * and ** indicate statistical significance at the 10% and 5% levels. Standard errors clustered by school.*

Figure 5: Quantile regression estimates of the effect of the Mathematics Mastery intervention across the non-calculator test score distribution



__Notes:__ Authors' calculations. Missing dummy variables included where information on covariates is not available. Dashed horizontal line illustrates the OLS estimate. Solid black (grey) circular markers indicate whether the intervention effect is significantly greater than 0 at the 5% (10%) level. Standard errors bootstrapped by cluster using 50 replications. Dependent variable is children's score on the non-calculator section of the PiM test. A full set of parameter estimates can be found in Appendix B.

As previously noted, around 40% of PiM test questions covered material *not* part of the Year 7 Mathematics Mastery curriculum. A clear advantage of this test was that it was therefore not too closely aligned to the Mathematics Mastery intervention, and hence there was less risk of the intervention group having been 'taught to the test'. However, it is also possible that this could have depressed the effect picked up by the test. To investigate this issue, and investigate the possibility at potential substitution effects, two sub-scales were created within the PiM test. One was formed of test questions covered by the Mathematics Mastery curriculum (60% of all test questions asked), with the other formed of questions that were not (the remaining 40% of the test). (These questions were chosen by Ark blind to our analysis.) It was expected that children in the intervention group would do no better (and possibly worse) on questions covering material that was not part of the Mathematics Mastery curriculum. In contrast, a positive intervention effect was expected on questions where the Mathematics Mastery programme places more time, effort, and emphasis.

Table 12 presents the results. As perhaps expected, the Mathematics Mastery intervention did not have any impact upon children's performance on questions covering topics outside the Mathematics Mastery curriculum. Indeed, the estimated intervention effect is essentially zero (effect size = –0.003). In contrast, the intervention had a more pronounced effect upon material that was focused upon within the Mathematics Mastery curriculum (effect size = 0.100), just reaching statistical significance at the 5% level (t = 2.15; p = 0.04). Interestingly, a similar effect size is found for the FSM-only sample (effect size = 0.104), with the confidence interval ranging between 0 and 0.21 standard deviations.

Table 12: The impact of the Mathematics Mastery secondary school programme on children's 'Progress in Maths' test scores

	Not covered in MM		Covered in MM	
	Beta	**SE**	**Beta**	**SE**
Intervention group (Ref: Control)				
Intervention	**–0.003**	**0.041**	**0.100****	**0.047**
Key Stage 1 mathematics (Ref: Level 1)				
Level 2a	**0.208****	**0.054**	**0.222****	**0.049**
Level 2b	**0.161****	**0.045**	**0.130****	**0.041**
Level 2c	**0.077****	**0.037**	**0.046**	**0.034**
Level 3	**0.305****	**0.066**	**0.351****	**0.064**
Key Stage 1 average points score	**0.092****	**0.023**	**0.080****	**0.021**
Key Stage 2 maths test score	**0.599****	**0.019**	**0.667****	**0.018**
Key Stage 2 maths test score squared	**0.080****	**0.009**	**0.139****	**0.010**
Key Stage 2 English score	**0.107****	**0.015**	**0.106****	**0.015**
FSM (Ref: No)				
Yes	**–0.105****	**0.021**	**–0.092****	**0.021**
Ethnic group (Ref: Other)				
Asian	**–0.094***	**0.054**	**-0.070**	**0.071**
Black	**–0.200****	**0.043**	**–0.170****	**0.060**
Chinese	**0.196**	**0.142**	**0.167***	**0.091**
Mixed	**0.021**	**0.060**	**–0.058**	**0.063**
Unclassified	**0.078**	**0.116**	**0.059**	**0.091**
White	**–0.041**	**0.048**	**–0.103***	**0.061**
Gender (Ref: Female)				
Male	**–0.122****	**0.023**	**–0.090****	**0.022**
English as additional language (Ref: No)				
Yes	**0.067****	**0.029**	**0.024**	**0.029**
Constant	**–0.092**	**0.059**	**–0.180****	**0.074**
N	**5,888**		**5,884**	

Notes: *Authors' calculations. Intervention effect presented in the columns labelled 'Beta'. SE stands for standard error. * and ** indicate statistical significance at the 10% and 5% levels. Standard errors are clustered at the school level.*

Robustness tests

Multiple Imputation by Chained Equations (MICE) has been used to ensure observations are not dropped from the analysis where covariate information is missing. Table 13 considers the robustness of results to alternative ways of handling missing covariate data. The left-hand columns ('multiple imputation') present results from the main analysis (as per Table 9). In the middle two columns, 'missing' dummy variables are added to the model—thus maintaining sample size without having to impute data. The final set of columns provide the 'complete case' analysis, where observations are excluded from the model if information is missing on any variable (i.e. list-wise deletion).

Reassuringly, the estimated intervention effect is robust to these various different ways of handling missing covariate data. The effect size ranges from +0.055 in the multiple imputation analysis to +0.061 in the complete case analysis. On no occasion does the intervention effect reach statistical significance at the 5% level.

Table 13: A comparison of results across different methods for handling missing covariate data

	Multiple imputation		Missing dummy		Complete case	
	Beta	SE	Beta	SE	Beta	SE
Intervention group (Ref: Control)						
Intervention	0.055	0.046	0.059	0.046	0.061	0.046
Key Stage 1 maths (Ref: Level 1)						
Level 2a	0.21**	0.05	0.34**	0.06	0.36**	0.06
Level 2b	0.14**	0.04	0.23**	0.05	0.24**	0.05
Level 2c	0.06*	0.03	0.11**	0.04	0.11**	0.04
Level 3	0.32**	0.06	0.51**	0.08	0.53**	0.08
Key Stage 1 average points score	0.09**	0.02	0.05**	0.02	0.06**	0.02
Key Stage 2 maths test score	0.66**	0.02	0.63**	0.02	0.62**	0.02
Key Stage 2 maths test score squared	0.12**	0.01	0.12**	0.01	0.12**	0.01
Key Stage 2 English score	0.11**	0.01	0.11**	0.01	0.11**	0.01
FSM (Ref: No)						
Yes	–0.10**	0.02	–0.11**	0.02	–0.11**	0.02
Ethnic group (Ref: Other)						
Asian	–0.08	0.06	–0.08	0.06	–0.09	0.07
Black	–0.19**	0.05	–0.18**	0.05	–0.19**	0.06
Chinese	0.00	0.21	–0.06	0.23	–0.11	0.24
Mixed	–0.03	0.06	–0.02	0.06	–0.03	0.07
Unclassified	0.08	0.09	0.04	0.09	0.09	0.10
White	-0.08	0.05	–0.07	0.05	–0.10	0.06
Gender (Ref: Female)						
Male	–0.11**	0.02	–0.12**	0.02	–0.11**	0.02
English as additional language (Ref: No)						
Yes	0.05**	0.03	0.06**	0.03	0.06**	0.03
Constant	–0.15**	0.06	–0.24**	0.07	–0.23**	0.07
n	5,919		5,616		5,175	

Notes: *Authors' calculations. Dependent variable is total PiM score (standardised to mean 0 and standard deviation of 1). SE stands for standard error. * and ** indicate statistical significance at the 10% and 5% levels. Standard errors are clustered by school.*

External validity

Schools were not randomly selected for the trial. Rather, Ark, which was running the intervention, was allowed to choose schools. In this section we compare the characteristics of children participating in the trial to the state school population in England.

Table 14 considers whether pupils within the 50 initially randomised schools have similar baseline (KS1 and KS2) test scores to pupils in the rest of England. Trial participants, on average, performed less well in their KS1 and KS2 examinations than the state school population as a whole. For instance, their KS1 average points scores (and KS2 maths test scores) were approximately 0.2 standard deviations (0.1 standard deviations) below the population mean. This seems to be driven, at least in part, by the fact that the trial particularly under-represented high achievers (relative to the population). For instance, just 12% of children participating in the trial were awarded Level 3 in their Key Stage 1 maths test, compared to 19% of all state school pupils in England.

Table 15 presents a similar comparison in terms of other observable characteristics. Of children enrolled in the trial, 29% were eligible for FSM, compared to 18% of pupils in the population. This suggests that trial participants were much more likely to come from a low-income background. Similarly, ethnic minorities were over-represented in the trial—particularly Black (19% in the sample versus 5% in the population) and Asian (20% in the sample versus 10% in the population) groups.

Table 14: A comparison of Key Stage 1 and Key Stage 2 test scores of trial participants to the England state school population

	Trial participants	England
Key Stage 1 maths		
Level 1 %	**11**	**8**
Level 2A %	**24**	**25**
Level 2B %	**24**	**24**
Level 2C %	**18**	**15**
Level 3 %	**12**	**19**
Missing %	**12**	**8**
Key Stage 1 reading		
Level 1 %	**15**	**12**
Level 2A %	**20**	**22**
Level 2B %	**22**	**21**
Level 2C %	**13**	**12**
Level 3 %	**16**	**24**
Missing %	**13**	**9**
Key Stage 1 writing		
Level 1 %	**18**	**14**
Level 2A %	**15**	**18**
Level 2B %	**24**	**26**
Level 2C %	**22**	**20**
Level 3 %	**7**	**11**
Missing %	**14**	**10**
Mean (SD) KS1 total points score	**14.6 (3.5)**	**15.3 (3.6)**
Mean (SD) KS2 maths score	**68.4 (20.9)**	**70 (21)**
Mean (SD) KS2 reading score	**31 (10.2)**	**33 (10)**
School n	**7,224**	**3,996**
Pupil n	**50**	**531,145**

Notes: *Authors' calculations using the NPD. Figures for England refer to state school pupils only.*

Table 15: A comparison of demographic characteristics of trial participants to the England state school population

	Trial participants	England
Eligible for FSM		
No %	**71**	**82**
Yes %	**29**	**18**
Gender		
Female %	**48**	**49**
Male %	**52**	**51**
Ethnic group		
White %	**49**	**78**
Asian %	**20**	**10**
Black %	**19**	**5**
Mixed %	**7**	**5**
Chinese %	**0**	**0**
Other/unclassified %	**5**	**2**
School n	**50**	**3,996**
Pupil n	**7,712**	**531,145**

Notes: *Authors' calculations using the NPD. Figures for England refer to state school pupils only.*

Cost

There is an upfront cost of £6,000 for participating in the first year of the programme. (This is an 'at cost' price charged by Ark to cover basic infrastructure.) Ten days of staff time is required for training: half a day for the headteacher, two and a half days for the head of maths, and one day for each maths teacher (there were on average seven maths teachers per school). To calculate the cost of headteachers' time, we take the median point on the headteacher pay scale in England and Wales (£75,222[11]). This is then divided by 230 (the approximate number of working days in a year) to give a headteacher day rate of £327.[12] We then inflate this figure by a fifth to allow for other costs not directly incorporated into headteachers' salaries (e.g. employer contributions to pensions), giving a total cost of £392.[13] Analogous calculations have been made for heads of mathematics (two and a half days training at a final day rate of £251)[14] and class teachers (a total of seven days training, one day per teacher, at a final day rate of £141).[15] Thus total training costs are therefore equal to £1,740 per school per annum. This gives a total cost per secondary school of £7,740. There were, on average, 154 pupils per secondary school (7,712 children across the 50 initially recruited schools). The per-pupil cost therefore equals £50 per annum for this first year, with the cost per pupil likely to reduce in future years.

[11] This information has been drawn from http://www.education.gov.uk/get-into-teaching/about-teaching/salary/pay-and-benefits.
[12] The headteacher pay scale in England and Wales (outside London) ranges from £43,232 to £107,210. We have assumed headteachers work 46 five-day weeks per year (with the other six weeks as holiday).
[13] We appreciate that this is a rather crude way of accounting for such additional costs. However, using a substantially higher or lower figure here does not radically alter our results.
[14] We have assumed the head of maths to be on the 'leading practitioner' pay scale, which ranges from £38,215 to £58,096 (median £48,155).
[15] It is assumed the teachers will be on the 'main' pay scale, which ranges from £22,023 to £32,187 (median £27,105).

Process evaluation results

As explained in the description of the methodology, the findings as reported in this section of the report are based on school teacher perceptions as identified during the focus group discussions and classroom observations in each of those schools, the telephone interviews with two teachers from each of these five schools, and the questionnaires completed by many teachers from nearly all the schools.

Implementation

The tension between trying to ensure that pupils had a deep conceptual understanding of the topic and that they developed good fluency in procedures and techniques was a factor throughout the implementation. In many ways, this goes to the heart of the Mathematics Mastery approach, which attempts to shift the balance towards conceptual depth, and so it is not surprising that this created difficulties for some teachers. This tension was observed in a number of different ways:

- Many teachers found it difficult to adapt to an approach where the emphasis was on differentiation by outcome,[16] with little or no additional explicit support for lower achievers [see also 'Outcomes'].

- Concern was expressed in different forms about the nature of the supplied materials. One school [C] argued in the focus group discussions that the materials did not offer sufficient support for reinforcement, though one teacher in this school disagreed, believing that the materials were ideal preparation for proof. It became apparent in the focus group discussions that all the focus schools needed to revise the materials in one way or another, because the school perceived the need either to provide additional skills practice through written exercises or to better tune the materials for differing abilities.

- The problem-solving approach of the Mathematics Mastery intervention did not lend itself easily to typical homework practice in some schools. In fact, schools generally found the setting of homework problematic. Prior to the intervention, the focus schools had used a mixture of written assignments and online work for homework, with some schools emphasising the former and others stressing the latter. Teachers reported how they had to respond to the lack of homework tasks offered by Mathematics Mastery. In one case [B] at least, the response was in the light of parental concern. Teachers also reported that such a response was not easy. In another case [C], it was felt the lesson objectives were not easy to identify and so focused homework was not always possible. Some teachers took to writing specific worksheets. One school [A] felt there was a need for a Mathematics Mastery homework book. All five schools reflected this issue in the telephone interviews (it was mentioned by eight of the ten teachers interviewed), with all but one [C] mentioning they had needed to find separate homework tasks. As the programme develops, plans are in place to provide more structured 'intelligent practice' resources.

Teachers regularly made comments during the focus group discussions and the telephone interviews about the quality of the Mathematics Mastery materials, quite apart from the difficulty with differentiation mentioned above. These comments varied hugely, with some teachers welcoming the

[16] 'Differentiation by outcome' in mathematics education refers to a practice in which teachers give pupils tasks which have a range of possible outcomes, either in terms of the method used to reach an answer or in terms of there being multiple acceptable solutions. In some subjects this approach is de rigueur. For example, when an English teacher faced with differing abilities in his/her class asks pupils to write an essay under a title which the teacher supplies, the teacher knows that the pupils will differentiate themselves according to how they approach the task—that is, how they write the essay from the common starting point. In mathematics, it is generally regarded as more challenging to design tasks that are sufficiently open to allow such differentiation between abilities, because there is often only one correct or expected method and solution to a mathematics task. Enquiry-based tasks, such as investigations and practical activities, are sometimes used by mathematics teachers to differentiate by outcome.

open style of the tasks, especially the 'Do It Now' activities, but others criticising the amount of work they had to do in terms of downloading, reproducing, organising, and identifying the learning objectives. Our sense was that, while some of these problems might have been avoided, they became a focus of negativity when the teacher concerned did not feel aligned with the Mathematics Mastery philosophy.

- Education is a long-term process and teachers are used to having a sense of development across several years. One of the difficulties with interventions is that teachers can lose sight of—or be unaware of—how the immediate changes will impact on the longer-term development of pupils. The Mathematics Mastery programme is planned as a five-year professional development programme and, although the curriculum plan for all of Key Stage 3 was shared with schools from the start, in the focus group discussion at one school [B], teachers reported concerns that they were unaware which aspects of maths would be revisited in future years and whether the materials would be suitable for Year 9. This concern was echoed in one telephone interview [C].

- The success of the intervention seemed to depend largely on the prior beliefs of the teachers. Just one year into the intervention, we are unlikely to see a significant impact on deep-seated beliefs. School [C] was particularly non-amenable to the intervention. The leader of Mathematics Mastery in the school was a strong supporter of the approach, but most of the teachers in the mathematics department seemed to believe in teaching through practising skills and did not adjust well to the Mathematics Mastery pedagogy. Although in the perception of the process evaluation team there were positive findings in this school, it seems unlikely that any progress will be sustained. There was also resistance in two other schools [B and D]. It was pointed out in a telephone interview with one school [A] that teachers following the Mathematics Mastery programme need to have really strong questioning skills.

Fidelity

At the outset of the intervention, teachers tended to feel they had to follow the materials very closely. Over the course of the year, it seems that teachers in most of the focus schools became more confident about their ability to adapt the Mathematics Mastery lessons and resources or to substitute their own resources. Such adaptation was expected by the Mathematics Mastery project team and gradually teachers became more aware of this expectation. Teachers in two schools [A and B] noted that they adapted the expected timing and structure of lessons, sometimes spending much longer than intended on the 'Do It Now' activity. Teachers in another school [C] had at times "stepped away from" the Mathematics Mastery materials in order to use alternative approaches to achieve the intended learning outcomes in the lesson. Such adaptation would only be a threat to fidelity if schools were abandoning the Mathematics Mastery principles. Generally, teachers were trying to follow the Mathematics Mastery approach even when they chose to use alternative materials—though doing so caused tensions in some schools.

- In spite of the fact that some teachers expressed discontent with the principle of spending more time on fewer topics, all had followed the intended timing and order of the scheme of work. This was a particular challenge in one school [A], which had needed to fit the proposed five hours of teaching per week into three timetabled hours.

- One school [A] decided to implement the programme this year using younger teachers in the department in the intervention. It was felt that they would be more likely to adapt to the philosophy than the older teachers, who had more personal investment in conventional approaches. The intervention appeared to be quite successful in this school but there may be problems when it is rolled out to the other teachers in the mathematics department.

- Two schools [C and D] dropped using the pre- and post-tests during the year, reacting to what they perceived to be over-assessment.

Another school [A] observed in a telephone interview that the test results did not show as much progress as expected from the class work. The suggestion was that the learning witnessed in class time was different in nature. Whereas previously the pupils could answer questions without understanding, now they really had a proper grasp of the topics. As a result, they predicted more progress in later years than was showing in the current pre- and post-tests. We detected that the lack of progress in the short term could pose a threat to continued support for the intervention from senior management in some schools.

One school [B] did not provide formal collaboration time for the teachers to meet and plan, though they intend to do so next year.

Outcomes

One of the major objectives for the Mathematics Mastery project was for teachers to build fluency and understanding of underlying mathematical concepts.

There was evidence of success in this respect:

There was evidence of a shift during the intervention away from an emphasis on teaching procedures without an underpinning understanding towards an expectation that pupils should offer explanations and justifications. In some cases this was accompanied by a perceived increase in the level of independence among the pupils. In three of the five focus schools [A, B and C], teachers had originally perceived an emphasis on procedural approaches. Teachers in two of these schools [A and B] suggested that their approach had changed radically, with much more emphasis on the conceptual, expecting pupils to give explanations. Some teachers in one school [C] seemed concerned that the balance was too much towards the conceptual, although at least one teacher took the opposite view. In a fourth school [D], the reason for teaching mathematics tended to be explained in terms of its real-world utility rather than in conceptual terms. There was limited evidence that there was a shift during the intervention, with a greater expectation that answers should be justified and ideas constantly challenged by the teacher. In the fifth school [E], prior to the intervention, there was no evidence of an emphasis on either the procedural or the conceptual and after the intervention there was still a mixed response from teachers. The telephone interviews in three schools [A, B and C] reinforced this picture but at the same time they exposed teacher concerns that there may now be too little emphasis on practice, which may suggest that the undoubted change in practice was not always reflective of a change in beliefs. Such an interpretation would not be surprising as it is widely understood in the research literature how difficult it is to change deep-seated beliefs.

Four of the five focus schools [A, B, D, and E] reported a positive response to problem-solving tasks, with two of them [A and B] claiming that pupils were more willing to persist with problem-solving, demonstrating ‘resilience’. Teachers in one school [E] reported more open and exploratory work, but those in another [D] had been advised (it is not clear by whom, as this is not a message advocated by Mathematics Mastery) to omit investigation tasks for lower-attaining groups as these were too difficult. The telephone interviews in four schools [B, C, D, and E] provided further evidence of a positive response to problem-solving.

More than half of the 27 teachers from intervention schools who completed both questionnaires reported increases in the regularity of introducing concepts or skills through problem-solving and pupils working on unfamiliar problems, and less regularity in practice exercises. Nearly all of the other 27 teachers reported no change. These findings were supported by analysis of the whole set of questionnaires for the teachers in intervention and control schools, except there was not strong evidence in support of the finding that practice exercises had reduced in regularity. It does appear, though, that we can conclude with some confidence that teachers in the intervention had increased the regularity of an enquiry-based approach more than had teachers in the control schools. However, these changes may reflect a response to the requirements of the intervention rather than changes in the teachers' underlying beliefs. The questionnaire found little change in these teachers' views about the relationship between teaching concepts and practising procedures.

The Mathematics Mastery programme aimed to make more use of manipulatives to support conceptual understanding. All the focus schools reported increased use of practical work with manipulatives as a result of the intervention. However, according to focus group discussions and telephone interviews in three schools [A, B and E], the extent of use decreased later in the year. Reasons for this decrease were given variously as related to the complexity of the mathematical topics [B] or because pupils became 'embarrassed' about using them [E]. There was also a suggestion [B] that, as teachers became more confident in their ability to choose between resources and to alter the resources offered by the intervention, they may have chosen not to use manipulatives, especially with higher-attaining classes or for behavioural reasons with lower abilities. More often, however, the manipulatives were seen to be especially effective with lower-attaining groups.

Similarly, the use of diagrams and other visual aids was seen by the Mathematics Mastery project as central to their approach to support better understanding. Prior to the intervention, the use of diagrams to support learning and problem-solving varied considerably between the schools, with only one school [E] reporting extensive use of visual tools. The use of bar model diagrams was received very positively by teachers in all five focus schools [A, B, C and E in the focus group discussions and D in the telephone interview]. In two schools [B and C], some teachers had adopted these for use with further topics in older year groups. Teachers in another school [A] considered them to be especially useful for lower-attainers, enabling them to "access Level 4" work with fractions and algebra. The use of bar models was mentioned by many teachers as new to them and very effective.

The increased use of manipulatives and diagrams may reflect a response to the requirements of the intervention rather than an underlying change in beliefs or attitudes by teachers in the intervention schools. There appeared to be little short-term change in attitude among the 27 teachers from intervention schools who completed both questionnaires about the significance of using manipulatives and diagrams when teaching topics such as fractions or place value. Neither did there appear to be a change in beliefs about the benefits of drawing on real-life contexts when teaching percentages.

In our process evaluation, we wondered whether there would be a greater emphasis on trying to support deeper understanding through paired or group work, with less emphasis on the teacher talking from the front of the class. In fact, the amount of teacher talk had been identified as a concern by several of the focus schools before the intervention, and a high proportion of teacher talk was identified in the majority of baseline observation lessons. Following the intervention, four of the schools [A, B, C, and E] reported increased use of discussion, though the evidence of the observations suggests that in most cases this is likely to include more teacher–pupil discussion than pupil–pupil talk. Teachers in one school [C] reported an *increase* in teacher talk, needing to explain tasks to pupils. Concern was expressed in three of the focus schools [A, B, and D] about a lack of individual work involving 'work in their books', with one school [D] suggesting that this might be poor preparation for examinations. The shift to a more open approach does not seem to have been accompanied by reduced teacher talk, though the content of the teacher talk may well have become less prescriptive. The telephone interviews confirmed this analysis since four schools [A, B, C, and E] reported increased pupil discussion, though this was variously described as 'teacher led', 'teacher facilitated', or 'teacher questioning'. With the exception of one school [D], which reported less teacher talk, we were left with the impression that there was a significantly increased level of teacher–pupil interaction, which was regarded as discussion, and rather less time spent working on exercises. We had no evidence to suggest there was either increased or decreased pupil–pupil discussion, except in the telephone interview with one teacher [A], in which it was reported that pupils felt more comfortable getting help from peers.

The Mathematics Mastery approach of working together as a class on relatively open tasks requires, at least for some of the time, that teachers adopt methods of differentiating by outcome (see footnote 16). One school [D] was able to develop its work on methods of differentiation by outcome using the Mathematics Mastery materials. However, there was a concern expressed during both the focus group discussions and the telephone interviews in four of the five focus schools [A, B, C, and E] that the materials were not well suited to the lowest-attaining pupils. As a result, these schools had needed to adapt the materials considerably for such ability groups. One teacher [C] suggested that the gap between the highest and lowest attainers had widened. Another school [E] also commented that the materials were also not sufficiently challenging for the highest-attaining children, who were frustrated by revisiting at length the same topics they had already encountered at primary school. Although this observation was also made in other schools, it was generally felt that the children gradually began to realise that they were in fact enjoying the subject more by gaining extra understanding.

In fact, one might expect that pupils would enjoy having to seek out mathematical methods and solutions. We found evidence of higher levels of pupil enjoyment of mathematics and better explanations by the pupils. However, pupil responses did seem to vary across the focus schools. In two schools [A and E], teachers reported in the focus group discussions and telephone interviews higher levels of enjoyment, more precise explanations of the mathematics, and a better appreciation of what counts as a mathematical explanation by the pupils. In one school's telephone interviews [E], a teacher reported how Year 7 pupils had said in a survey that they really enjoyed maths. Similarly, in another school's telephone interviews [A], one teacher reported that 8.5 out of 10 pupils now said they enjoyed maths, whereas prior to the intervention it had been 4 or 5 out of 10. Another teacher in the same school reported that pupils were much more engaged than Year 7s had been in that teacher's previous teaching experience. A third school [B] was similarly positive about the impact on higher-attaining children but a 'tighter' teaching approach had been adopted with the lower-attaining sets. In their telephone interviews, this school was similarly positive about the intervention's impact on pupils. The other two schools reported no significant change. There was one exception to this generally positive picture [C] where many of the teachers, apart from the leader of Mathematics Mastery in the school, disliked the scheme, and the leader felt that this dislike had probably transferred from the teachers to the pupils.

The time for planning and collaboration that was made possible by participation in the project was received differently in the five focus schools. Teachers in one school [E], who had shared planning and resources prior to the intervention, and in another school [B], where there was no evidence of collaboration prior to the intervention, welcomed the weekly meeting time, even though in the latter case this had not been formally timetabled by the school. However, in one school [C] teachers perceived the meetings as not useful because classes with different needs required separate planning.

The analysis of the questionnaires appears to support the finding for increased collaboration among teachers in the intervention schools. Many of the 27 teachers from intervention schools who completed both questionnaires reported increases in the regularity of: planning with colleagues, discussing the teaching of particular topics and mathematical ideas. Around three-quarters of the 27 teachers reported such changes. Nearly all of the others reported no change. Analysis of the questionnaires for the complete set of intervention and control teachers also showed that there was evidence that intervention teachers changed their practice towards increased collaboration through extra shared planning, and discussion of topics and mathematical ideas. There is therefore reason to be confident about this finding.

Even though it was not perhaps an explicit aim of the Mathematics Mastery project, we wondered in our process evaluation whether, in a more open and discussion-based situation, teachers might feel more able to encourage connections across apparently different parts of the mathematics curriculum and with practice outside the classroom. We cannot conclude that this was not the case on the basis of our discussions and observations, but we found very little evidence of it.

Formative findings

- Some evidence from one school [C] suggests that there may be a need to consider materials that offer reinforcement of ideas. However, there is clearly a tension between adopting the Mathematics Mastery approach and giving numerous exercises for practice. The comment may reflect a different pedagogic position in that school from that of the project, with the result that some teachers expressed this concern (see the section on Implementation).

- The intervention should consider how it might give more advice and support with respect to homework setting.

- The intervention should consider how it might give more advice and support with respect to differentiation.

- The intervention should consider how it might give more advice and support with respect to a teacher's skills in questioning and facilitating pupil–pupil discussion.

- Some evidence from school B would suggest that there is a need to ensure that teachers implementing the scheme are aware of how the current work relates to that which the pupils will meet in later years.

- We examined the findings above in relation to the characterisations of the schools in terms of size, number of pupils in receipt of free school meals and previous progress with respect to mathematics achievement by the pupils. Overall, schools A, D, and E perhaps reported the greatest success but there is no clear pattern in how the profiles of these three schools differ from those of schools B and C. Informally, a much more pertinent factor seemed to be the prior pedagogic beliefs of the teachers in relation to the objectives of the Mathematics Mastery project. The mathematics departments in schools B and C seemed to include several teachers whose beliefs about teaching were not consistent with those of Mathematics Mastery and whose beliefs appeared not to be changed by one year's implementation of the intervention.

Control group activity

The only data we have about activity in the control schools is found in the analysis of the questionnaires at the start and end of the intervention. The questionnaire data essentially supports the notion that business was as usual in the control schools. None of the questions showed any significant change when the responses at the end of the intervention were compared with those at the start.

Conclusion

Key conclusions
1. On average, Year 7 pupils in schools adopting Mathematics Mastery made a small amount more progress than pupils in schools that did not. However, the effect detected was not statistically significant, meaning that it is not possible to rule out chance as an explanation.
2. There is no strong evidence that the approach had a greater impact on lower-attaining pupils than on higher-attaining pupils.
3. Combining the findings from this study and a second randomised controlled trial of Mathematics Mastery involving Year 1 pupils may strengthen the overall evidence for the approach.
4. Given the low per-pupil cost, Mathematics Mastery may represent a cost-effective change for schools to consider. However, teachers would need to resolve tensions related to differentiation to provide support for all groups of children.
5. It would be worthwhile to track the medium and long-term impact of the approach, to assess whether there is a cumulative effect to the approach and whether it has an impact on performance in high-stakes tests.

Limitations

The findings outlined above should be considered within the context of the limitations of this study. The following factors particularly stand out.

1. **Short-term follow-up only**: This evaluation assessed the impact of Mathematics Mastery on pupils in Year 7, after the programme had been implemented in schools for one year. It is not possible to extrapolate these findings to longer-term outcomes, such as performance in GCSE exams.

2. **Small 'dose' of the intervention**: Children have been exposed to the Mathematics Mastery intervention for just one academic year. This may be considered a relatively small 'dose' of the programme, as children will actually follow the Mathematics Mastery curriculum throughout their time at secondary school. Yet little is known about the possible cumulative impact of the programme, when children have been exposed to Mathematics Mastery over a prolonged period of time.

3. **A new programme in schools**: The Mathematics Mastery intervention introduced a new approach to teaching maths within intervention schools. Teachers within these schools would have had to get used to the Mathematics Mastery curriculum and how to teach it effectively. In contrast, teachers in control schools were to proceed with 'business as usual'—teaching the same material as they have in previous years. It is possible that teachers in the intervention schools may become more effective at implementing the Mathematics Mastery programme as they gain more experience using it.

4. **External validity:** Participating schools volunteered to take part in the project, so it is not possible to say whether similar effects would be seen in all schools.

5. **Low-stakes tests:** The post-test used in this trial was a 'low-stakes' assessment, meaning that neither participating children nor their schools had anything riding on the results. It has

been suggested that this may lead to a lack of effort among children taking the test (Wise and DeMars, 2005). Moreover, this factor could potentially differ between intervention and control groups. For instance, if intervention schools are more engaged in the trial than control schools, they may have put more effort into completing the post-test. This would potentially lead to an upwards bias in the estimated intervention effect.

6. **Process data**: The findings in the process evaluation are largely based on teachers' perceptions, with very limited corroboration by classroom observations. While many of the issues reported are possibly best captured in this way, teachers' perspectives do not necessarily transform faithfully into pupil achievement.

Interpretation

The hypothesis tested in this study was that a one-year dose of the Mathematics Mastery programme would lead to a significant improvement in Year 7 (age 11/12) children's maths test scores. Although point estimates were consistent with a small, positive gain, the study did not have sufficient statistical power to rule out chance as an explanation. Within the context of the wider educational literature, the effect size reported (less than 0.10 standard deviations) would typically be considered 'small'. Indeed, it is well below the values of 0.5 to 0.6 reported in the meta-analyses conducted by Guskey and Piggott (1988), Kulik, Kulik, and Bangert-Drowns (1990) and Waxman et.al. (1985).

Yet, despite the modest and statistically insignificant effect, the Mathematics Mastery intervention has shown some promise. First, in a sister trial investigating the impact of Mathematics Mastery on a younger cohort of children (age 5/6), a similar effect size has been found (approximately 0.10 standard deviations for a one-year dose; Jerrim and Vignoles, 2015). Meta-analysing the findings from the primary and secondary studies shows a statistically significant positive impact overall. It is important to note the limitations of this meta-analytic approach, and the care needed in interpreting findings based on studies that may vary in important ways. However, given the similarity of the studies and the whole-school nature of the programme, this approach does appear to strengthen the evidence that the first year of the Mathematics Mastery programme does indeed lead to a small improvement in children's maths test scores. Second, although effect sizes tend to be small, so are the costs per pupil. This suggests that the programme may be cost-effective, and could actually be a wise investment for schools.[17] Third, children are likely to follow the Mathematics Mastery programme for a number of years (e.g. throughout secondary school), whereas this evaluation has only considered the impact of a one-year 'dose'. Long-term effects after sustained exposure to the programme could be significantly higher. Third, with any change to the curriculum, there is the potential for negative side-effects. This includes children suffering a decline in skills that no longer form part of the syllabus. The fact that there was no detectable impact on children's ability to use a calculator was thus a reassuring result.

Through process evaluation some challenges for implementation were identified. For example, some teachers felt unsure about the impact of the intervention on examination results, in particular in relation to the seemingly reduced content coverage in each year. In addition, some teachers were unsure how to differentiate between different ability groups when tasks were open and sometimes did not have explicit learning objectives. Given the low per-pupil cost, Mathematics Mastery may represent a cost-effective change for schools to consider. However, teachers would need to resolve tensions related to differentiation to provide support for all groups of children.

[17] Indeed, one may draw comparisons to The Literacy Hour – a change to the school curriculum in the late 1990's to raise basic English skills. In a quasi-experimental evaluation, Machin and McNally (2008) found an effect similar to that reported for the Mathematics Mastery programme here (0.08 standard deviations). Yet, due to the low per-pupil cost, the aforementioned authors reported the intervention to be 'cost effective' (Machin and McNally 2008:1458).

Future research and publications

There are four key questions for future research to address. First, this report has focused upon the impact of the first year of the Mathematics Mastery intervention, with outcomes measured immediately after the RCT has finished. Future work should investigate whether effect sizes increase, decrease, or remain stable after longer exposure to the programme, and whether any apparent effect persists over a longer period of time. Second, future evaluations might consider the impact of Mathematics Mastery when the schools and teachers involved have more experience of delivering this curriculum. Third, it should be noted that participating schools volunteered to take part in the project, so it is not possible to say whether similar effects would be seen in all schools. Were the approach to be scaled up in a different way, for example by mandating its use in certain schools, a further evaluation would be required. Finally, our study should be replicated using a high-stakes outcome test.

The project was one of two evaluations of Mathematics Mastery funded by the Education Endowment Foundation (EEF). A second project assessed the impact of Mathematics Mastery on pupils in Year 1. The evaluation report from this project, and an overall summary combining findings from both evaluations, are available on the EEF website.

A working paper documenting results from both trials (Jerrim and Vignoles, 2015) will also be available from http://ideas.repec.org/s/qss/dqsswp.html and www.johnjerrim.com in early 2015.

References

Boaler, J. (2010) *The Elephant in the Classroom: Helping Children Learn and Love Maths*. London: Souvenir Press.

Drury, H. (2014) *Mastering Mathematics: Teaching to transform achievement*. Oxford: Oxford University Press.

Dweck, C. (2006) *Mindset: The new psychology of success*. New York: Random House.

Freudenthal, H. (1968) 'How to teach mathematics so as to be useful'. *Educational Studies in Mathematics*, 1: 3–8.

Guskey, T. and Piggott, T. (1988) 'Research on group-based mastery learning programs: a meta-analysis'. *Journal of Educational Research*, 81(4): 197–216.

Heddens, J. (1986) 'Bridging the gap between the concrete and the abstract'. *Arithmetic Teacher,* 33: 14–17.

Hoyles, C. (1985) 'What is the point of group discussion in mathematics?' *Educational Studies in Mathematics*, 16(2): 205–214.

Kulik, C.-L., Kulik, J., and Bangert-Drowns, R. (1990) 'Effectiveness of mastery learning programs: a meta-analysis'. *Review of Educational Research*, 60(2): 265–299.

Lee, C. (1998) 'Discussion in a mathematics classroom: developing a teacher's awareness of the issues and characteristics'. Oxford: Centre for Research into Mathematics.

Machin, S. and McNally, S. (2008) 'The literacy hour'. *Journal of Public Economics,* 92: 1,441–1,462.

Mulford, W., Silins, H., and Leithwood, K. (2004) *Educational Leadership for Organisational Learning and Improved Student Outcomes. Studies in Educational Leadership*. Springer.

Organisation for Economic Cooperation and Development (OECD) (2013) Programme for International Student Assessment 2012: Results from PISA 2012. United Kingdom Country Note. Available at: www.oecd.org/unitedkingdom/PISA-2012-results-UK.pdf (accessed 28 January 2015).

Skemp, R. (1976) 'Relational understanding and instrumental understanding'. *Mathematics Teaching,* 77: 20–26.

Sowell, E. (1989) 'Effects of manipulative materials in mathematics instruction'. *Journal for Research in Mathematics Education*, 20: 498–505.

Jerrim, J and Vignoles, A. (2015) 'Mathematics Mastery Secondary School: A Cluster Randomised Controlled Trial.' EEF report.

Waxman, H., Wang, M., Anderson, K., and Walberg, H. (1985) 'Adaptive education and student outcomes: a quantitative synthesis'. *Journal of Educational Research*, 78(4): 228–236.

Wise, S. and DeMars, C. (2005) 'Low examinee effort in low-stakes assessment: problems and potential solutions'. *Educational Assessment*, 10(1): 1–17.

Appendix A: Balance on baseline characteristics at the school level

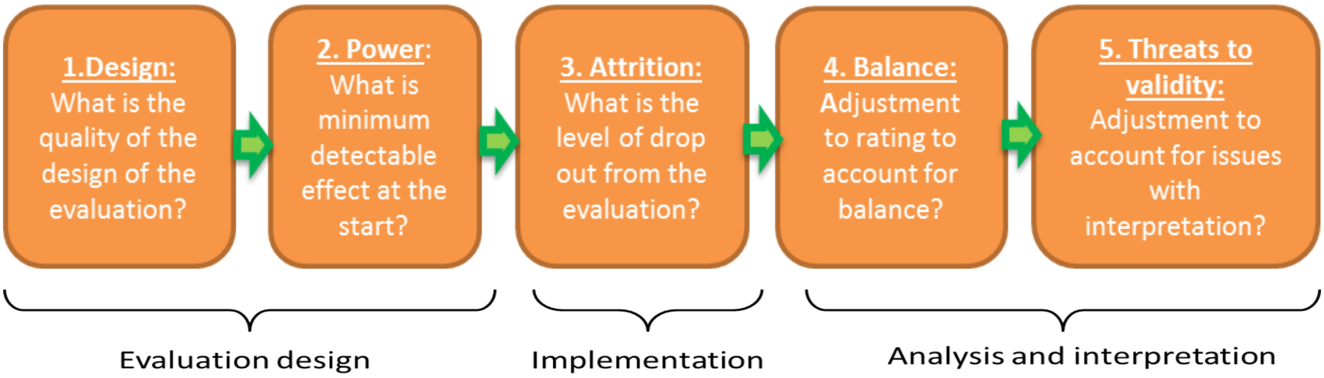
	As randomised		After attrition	
	Control	**Intervention**	**Control**	**Intervention**
KS1 APS	**–0.071**	**–0.033**	**–0.066**	**–0.011**
KS2 maths	**–0.031**	**–0.033**	**–0.031**	**–0.035**
KS2 English	**–0.068**	**–0.026**	**–0.053**	**–0.021**
% eligible for FSM	**28.2**	**30.3**	**29.1**	**28.9**
% female	**45**	**47.9**	**45.8**	**49.9**
School n	**25**	**25**	**21**	**23**

Appendix B. Quantile regression model parameter estimates

Percentile of the PiM non-calculator test distribution									
		Q15	Q20	Q25	Q30	Q35	Q40	Q45	Q50
Group (Ref: Control)									
Intervention	Beta	0.140	0.120	0.110	0.098	0.100	0.100	0.089	0.089
	SE	0.073	0.054	0.047	0.068	0.053	0.059	0.058	0.054
Key Stage 1 maths (Ref: Level 1)									
Level 2A	Beta	0.250	0.310	0.340	0.380	0.400	0.400	0.380	0.330
	SE	0.098	0.096	0.070	0.082	0.078	0.077	0.087	0.066
Level 2B	Beta	0.130	0.180	0.200	0.250	0.260	0.250	0.230	0.190
	SE	0.070	0.060	0.057	0.064	0.064	0.060	0.067	0.051
Level 2C	Beta	0.025	0.057	0.094	0.140	0.150	0.140	0.140	0.110
	SE	0.053	0.056	0.045	0.048	0.038	0.039	0.046	0.043
Level 3	Beta	0.480	0.550	0.580	0.590	0.610	0.600	0.530	0.470
	SE	0.130	0.110	0.100	0.100	0.100	0.100	0.110	0.095
Key Stage 1 average points score	Beta	0.075	0.044	0.036	0.028	0.017	0.009	0.025	0.036
	SE	0.043	0.039	0.040	0.031	0.029	0.036	0.033	0.029
Key Stage 2 maths score	Beta	0.600	0.620	0.630	0.630	0.640	0.650	0.660	0.660
	SE	0.026	0.019	0.016	0.022	0.023	0.020	0.022	0.020
Key Stage 2 maths score ^2	Beta	0.160	0.150	0.150	0.140	0.140	0.130	0.130	0.120
	SE	0.013	0.014	0.012	0.015	0.016	0.011	0.012	0.011
Key Stage 2 English score	Beta	0.120	0.130	0.130	0.130	0.130	0.130	0.120	0.120
	SE	0.020	0.019	0.021	0.017	0.019	0.021	0.023	0.022
FSM (Ref: No)									
Yes	Beta	–0.130	–0.140	–0.120	–0.110	–0.110	–0.120	–0.120	–0.110
	SE	0.028	0.028	0.032	0.027	0.029	0.029	0.030	0.031
Ethnicity (Ref: White)									
Other	Beta	0.015	0.063	0.110	0.081	0.090	0.120	0.110	0.110
	SE	0.077	0.090	0.068	0.094	0.097	0.100	0.084	0.074
Asian	Beta	0.035	0.060	0.048	0.049	0.040	0.007	0.001	–0.010
	SE	0.061	0.054	0.051	0.052	0.061	0.052	0.051	0.066
Black	Beta	–0.110	–0.110	–0.120	–0.120	–0.120	–0.130	–0.130	–0.110
	SE	0.050	0.037	0.041	0.046	0.048	0.047	0.043	0.050
Chinese	Beta	0.540	0.590	0.580	0.470	0.400	0.320	0.300	0.240
	SE	0.290	0.270	0.170	0.180	0.100	0.120	0.099	0.088
Mixed	Beta	0.084	0.085	0.110	0.097	0.086	0.057	0.072	0.076
	SE	0.058	0.059	0.063	0.054	0.055	0.066	0.060	0.060
Unclassified	Beta	0.100	0.190	0.130	0.080	0.075	0.055	0.083	0.130
	SE	0.120	0.086	0.100	0.130	0.092	0.120	0.120	0.150
Gender (Ref: Female)									
Male	Beta	–0.190	–0.170	–0.160	–0.150	–0.130	–0.140	–0.120	–0.110
	SE	0.034	0.022	0.022	0.030	0.030	0.032	0.029	0.034
Constant	Beta	–0.860	–0.780	–0.710	–0.640	–0.580	–0.470	–0.360	–0.260
	SE	0.087	0.066	0.060	0.066	0.078	0.064	0.080	0.073
N		**5,560**	**5,560**	**5,560**	**5,560**	**5,560**	**5,560**	**5,560**	**5,560**

		Percentile of the PiM non-calculator test distribution						
		Q55	**Q60**	**Q65**	**Q70**	**Q75**	**Q80**	**Q85**
Group (Ref: Control)								
Intervention	Beta	0.073	0.059	0.059	0.052	0.052	0.046	0.029
	SE	0.054	0.054	0.043	0.046	0.043	0.041	0.036
Key Stage 1 maths (Ref: Level 1)								
Level 2A	Beta	0.320	0.360	0.330	0.330	0.300	0.270	0.190
	SE	0.074	0.060	0.068	0.067	0.061	0.062	0.053
Level 2B	Beta	0.170	0.220	0.210	0.200	0.170	0.140	0.077
	SE	0.063	0.048	0.055	0.058	0.060	0.058	0.037
Level 2C	Beta	0.089	0.097	0.083	0.096	0.062	0.043	–0.018
	SE	0.055	0.041	0.053	0.049	0.050	0.044	0.042
Level 3	Beta	0.450	0.480	0.470	0.450	0.400	0.330	0.240
	SE	0.098	0.074	0.086	0.097	0.087	0.082	0.072
Key Stage 1 average points score	Beta	0.040	0.017	0.017	0.031	0.042	0.050	0.044
	SE	0.029	0.029	0.031	0.030	0.030	0.030	0.027
Key Stage 2 maths score	Beta	0.650	0.650	0.650	0.650	0.640	0.650	0.650
	SE	0.018	0.016	0.018	0.017	0.020	0.016	0.016
Key Stage 2 maths score ^2	Beta	0.110	0.110	0.098	0.089	0.080	0.069	0.055
	SE	0.010	0.009	0.010	0.008	0.008	0.010	0.008
Key Stage 2 English score	Beta	0.130	0.140	0.150	0.140	0.140	0.140	0.140
	SE	0.019	0.016	0.020	0.017	0.018	0.015	0.017
Free School Meals (Ref: No)								
Yes	Beta	–0.120	–0.120	–0.120	–0.100	–0.082	–0.058	–0.032
	SE	0.032	0.027	0.030	0.026	0.028	0.028	0.027
Ethnicity (Ref: White)								
Other	Beta	0.100	0.110	0.120	0.085	0.082	0.038	0.018
	SE	0.076	0.076	0.056	0.051	0.060	0.045	0.046
Asian	Beta	0.009	–0.017	–0.034	–0.029	–0.010	–0.013	–0.010
	SE	0.063	0.045	0.060	0.053	0.045	0.046	0.045
Black	Beta	–0.100	–0.120	–0.130	–0.130	–0.110	–0.120	–0.120
	SE	0.050	0.039	0.037	0.040	0.033	0.033	0.033
Chinese	Beta	0.230	0.160	0.110	0.180	0.130	0.050	0.120
	SE	0.091	0.110	0.120	0.130	0.130	0.140	0.140
Mixed	Beta	0.079	0.073	0.077	0.059	0.053	0.035	0.055
	SE	0.049	0.043	0.047	0.045	0.036	0.041	0.037
Unclassified	Beta	0.098	0.099	0.064	0.100	0.072	0.027	0.078
	SE	0.120	0.096	0.098	0.093	0.091	0.110	0.140
Gender (Ref: Female)								
Male	Beta	–0.110	–0.110	–0.096	–0.083	–0.082	–0.070	–0.079
	SE	0.028	0.032	0.031	0.027	0.029	0.025	0.027
Constant	Beta	–0.150	–0.088	0.005	0.085	0.190	0.320	0.500
	SE	0.074	0.069	0.060	0.067	0.072	0.060	0.054
N		**5,560**	**5,560**	**5,560**	**5,560**	**5,560**	**5,560**	**5,560**

Appendix C: Padlock rating



Rating	1. Design	2. Power (MDES)	3. Attrition	4. Balance	5. Threats to validity
5 🔒	Fair and clear experimental design (RCT)	< 0.2	< 10%	Well-balanced on observables	No threats to validity
4 🔒	Fair and clear experimental design (RCT, RDD)	< 0.3	< 20%		
3 🔒	Well-matched comparison (quasi-experiment)	< 0.4	< 30%		
2 🔒	Matched comparison (quasi-experiment)	< 0.5	< 40%		
1 🔒	Comparison group with poor or no matching	< 0.6	< 50%		
0 🔒	No comparator	> 0.6	> 50%	Imbalanced on observables	Significant threats

The final security rating for this trial is 4 🔒. This means that the conclusions have moderate to high security.

The trial was designed as an efficacy trial and could achieve a maximum of 5 🔒. This was well conducted randomised controlled trial, with few threats to internal validity and a large sample size (MDES 0.14). However, there were moderate levels of attrition (6 schools out of 50 dropping out – 12%. 4 of the 6 schools that dropped out were in the control arm). There was some imbalance in the ethnicity of the two arms, but not sufficient to lose a padlock. There were no other substantial threats warranting loss of security. Therefore, the overall security rating is 4 🔒.

Appendix D: Cost rating

Cost ratings are based on the approximate cost per pupil of implementing the intervention over one year. Cost ratings are awarded using the following criteria.

Cost	Description
£	*Very low:* less than £80 per pupil per year.
£ £	*Low:* up to about £170 per pupil per year.
£ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix E: A meta-analysis of the Mathematics Mastery primary school and secondary school Randomised Controlled Trials

Two Mathematics Mastery trials were conducted simultaneously. The 'primary school' trial introduced Mathematics Mastery to Year 1 pupils (5/6-year-olds) and took place over two academic years (September 2012 to August 2013, and September 2013 to August 2014). The 'secondary school' trial was conducted in the September 2013 to August 2014 academic year, with the focus on Year 7 pupils (11/12-year-olds). A summary of this is presented in below, along with a meta-analysis of the results (note that the primary and secondary school trials have been given equal weight).

The reported effect size is similar across the two trials (0.10 for primary school and 0.06 for secondary school) though neither individually reaches statistical significance at the 5% level. Precision is increased, however, when information is combined across the two. Indeed, the pooled effect size of 0.073 is just significantly different from zero at conventional thresholds. Overall, these results support the conclusion that even a one-year dose of the Mathematics Mastery intervention leads to a small (yet potentially cost-effective) improvement in children's maths test scores.

Appendix Table E1: Meta-analysis results

	Primary school	Secondary school	Meta (combined)
Number of schools	83	44	127
School response rate	92%	88%	–
Number of pupils	4,176	5,938	10,114
Pupil response rate	82%	77%	–
Effect size	0.099*	0.055	0.073**
Standard error	0.054	0.046	0.035
95% confidence interval	–0.009 to 0.207	–0.037 to 0.147	0.005 to 0.142

Notes: *Authors' calculations. Meta-analysis has been weighted by standard error. Overall test scores (pre-specified primary outcome) have been reported for both trials. Huber–White adjustments have been made to all standard errors to account for clustering at the school level. * and ** indicate statistical significance of effect sizes at the 10% and 5% levels respectively*

This document is available for download at **www.educationendowmentfoundation.org.uk**

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk