Mathematics Mastery

Primary Evaluation Report

February 2015

Independent evaluators:

Anna Vignoles
John Jerrim
Richard Cowan

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

This project has been independently evaluated by Dr John Jerrim (Institute of Education, University of London), Professor Anna Vignoles (University of Cambridge) and with assessment advice from Professor Richard Cowan (Institute of Education, University of London).[1]

John Jerrim's research interests include randomised controlled trials, access to higher education, intergenerational mobility, cross-national comparisons and educational inequalities. He has worked extensively with the Programme for International Student Assessment (PISA), and has recently returned from working at the OECD. John won the inaugural ESRC Early Career Outstanding Impact award and has just received an ESRC grant to study cross-national comparisons of educational attainment and social mobility.

Anna Vignoles is Professor of Education at the University of Cambridge. She has published widely on the impact of school resources on pupil achievement and on the socio-economic gap in pupil achievement. Her research interests include issues pertaining to equity in education, school choice, school efficiency and finance and the economic value of schooling. Anna is a Research Fellow at the Institute for Fiscal Studies and a Visiting Professor at the Institute of Education. Anna has advised numerous government departments, including the Department for Education and the Department of Business, Innovation and Skills. Anna is also the economist member of the NHS Pay Review Body.

Richard Cowan is Professor of Psychology of Education at the Institute of Education. His interests are in the contribution that psychology can make to understanding children's development and education and he specialises in the study of maths in the primary phase.

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

---

[1] Initial advice on the project was also provided by Dr Rebecca Allen (Institute of Education) and we are grateful for this. All errors and omissions remain our own.

Contents

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' 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 1, 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 2 in the second year of implementation, and continue until the approach was in place across the school. Eighty-three schools from London and the South East participated in the trial, with a total sample of 4,176 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 7. An overall summary combining findings from both evaluations is available on the EEF website.

Key conclusions	
1.	On average, Year 1 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 little evidence that the effect of Mathematics Mastery differs between children with different levels of prior achievement.
3.	Combining the findings from this study and a second randomised controlled trial of Mathematics Mastery involving Year 7 pupils strengthens the overall evidence for the approach.
4.	Given the low per-pupil cost, Mathematics Mastery may represent a cost-effective change for primary schools to consider.
5.	It would be worthwhile to track the medium and long-term impact of the approach, in part to assess the degree to which the test used in this evaluation is predictive of general mathematics attainment and 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 two months' additional progress. However, the effect was not statistically significant, meaning that it is not possible to determine that it did not occur by chance.

It is not possible to assess at this stage whether the programme had a stronger or weaker impact on pupils eligible for free school meals compared to their peers. There is little evidence that the effect of Mathematics Mastery differs between children with different levels of prior achievement.

However, in a follow-up study Key Stage 2 data will be used to evaluate long-term impact of the programme on different groups of pupils.

How secure is this finding?

Overall, the findings from this evaluation are judged to be of moderate 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 begin the programme in 2012 or 2013. 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 when 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 7 pupils shows a statistically significant average impact of one additional month's progress. This combined finding strengthens the overall evidence for the approach, and is discussed in further depth in a summary report on the EEF's website.

Ninety-two per cent of schools and 82% of pupils who initially enrolled in the trial were successfully followed through to completion. There was a notable average difference in the initial level of attainment of pupils who participated in the programme compared to those in the comparison group, however the evaluators attempted to account for this difference through statistical analysis. 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.

Some caution is also needed since the test used to measure the impact of the intervention focused on some but not all aspects of pupils' mathematical skills. The extent to which these test results can be generalised to mathematics attainment in the longer term is difficult to assess.

How much does it cost?

The cost of the approach is estimated to be approximately £7,460 in the first year for a two-form primary school, including teacher training costs. The average 'per pupil' cost of the intervention is therefore around £131 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
Mathematics Mastery vs. comparison	4,176 pupils (83 schools)	+0.10 (-0.01, +0.21)	+2 months	🔒🔒🔒🔓🔓	££
For more information about evidence ratings, see Appendix A.					

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 1 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.[2,3]

Typically, mastery approaches involve breaking 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 that 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 is 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 hurdle receive additional instruction, whilst those who succeed engage in enrichment activity that seeks to deepen their understanding of the same topic. This differs from the Mathematics Mastery programme in that 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 (many of the studies reviewed were conducted in the 1960's, 1970's and early 1980's)—classroom environments, teaching quality, pedagogy and social context may have changed considerably since that time. Second, some studies

---

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

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

included in the Mastery section 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 have taken 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. Thus 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 1 pupils' maths test scores.

Project team and roles

Richard Cowan: provided expert advice on the 'Number Knowledge' maths test.

John Jerrim: led the data analysis and wrote the final report.

Anna Vignoles (Principal Investigator): led the trial design, randomisation and organisation of testing; had responsibility for overall management of the project.

Ethical review

The evaluation was submitted to the Institute of Education ethics committee. Ethical approval was submitted on 23 January 2012. Consent to participate in the study was given by the school rather than the individual child or their parents. Testers were, however, given guidance that any child who was distressed by the test should not be forced to finish the test and that if a teacher felt strongly that a child should not sit a test, perhaps due to having special educational needs, then again the child would not be forced to do so. Testers noted when this occurred. Twelve children were omitted from the testing for these reasons.

Methodology

Trial design

A clustered randomised controlled trial (RCT) was used, with random allocation at school level. As the Mathematics Mastery programme involves a change to the maths curriculum within schools, randomisation at either pupil or class level was inappropriate. Randomisation occurred after school recruitment for both cohorts of schools (see below) but before the pre-test took place.

Two school cohorts were recruited into the trial (henceforth Cohort A and Cohort B). A total of 40 schools were recruited to participate in the trial during the September 2012 to August 2013 academic year (Cohort A). Twenty of the Cohort A schools were randomly allocated to treatment, with the remaining 20 schools allocated to control. All Year 1 pupils within the Cohort A treatment schools then received the Mathematics Mastery programme during the 2012/13 academic year. The control schools, on the other hand, were asked to proceed with 'business as usual.' At the end of the Cohort A trial period (July 2013),[4] Cohort A control schools became eligible to receive the Mathematics Mastery programme.

A further 50 schools were then recruited to participate in the trial in the 2013/14 academic year (Cohort B). Of these, 25 were randomly assigned to the treatment group and 25 were assigned to control. All Year 1 pupils within the Cohort B treatment schools then received the Mathematics Mastery programme during the 2013/14 academic year. The control schools were asked to proceed with 'business as usual'.

In total, the trial therefore included 90 schools (45 treatment and 45 control) spread over the two academic years. Note that all Year 1 pupils within the treatment schools received the Mathematics Mastery programme and were thus considered to be participants in the trial.

Eligibility

The charity delivering the intervention (Ark) was responsible for recruiting schools to participate in the trial. The main exclusion criteria were that (a) schools could not already be receiving the Mathematics Mastery programme, and (b) private (independent) schools were ineligible. Otherwise, Ark was free to recruit any primary school within England. Schools were therefore purposefully 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.

Intervention

Mathematics Mastery is a not-for-profit organisation working with primary and secondary schools across England, and 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 who lead the international tables, such as Shanghai and Singapore (see Guskey, 2010). A 'mastery curriculum'

---

[4] No outcome measures from Cohort A pupils were collected in the second year of the trial.

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 practice day-to-day 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 & 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 launch year:

Overview of the Mathematics Mastery Programme
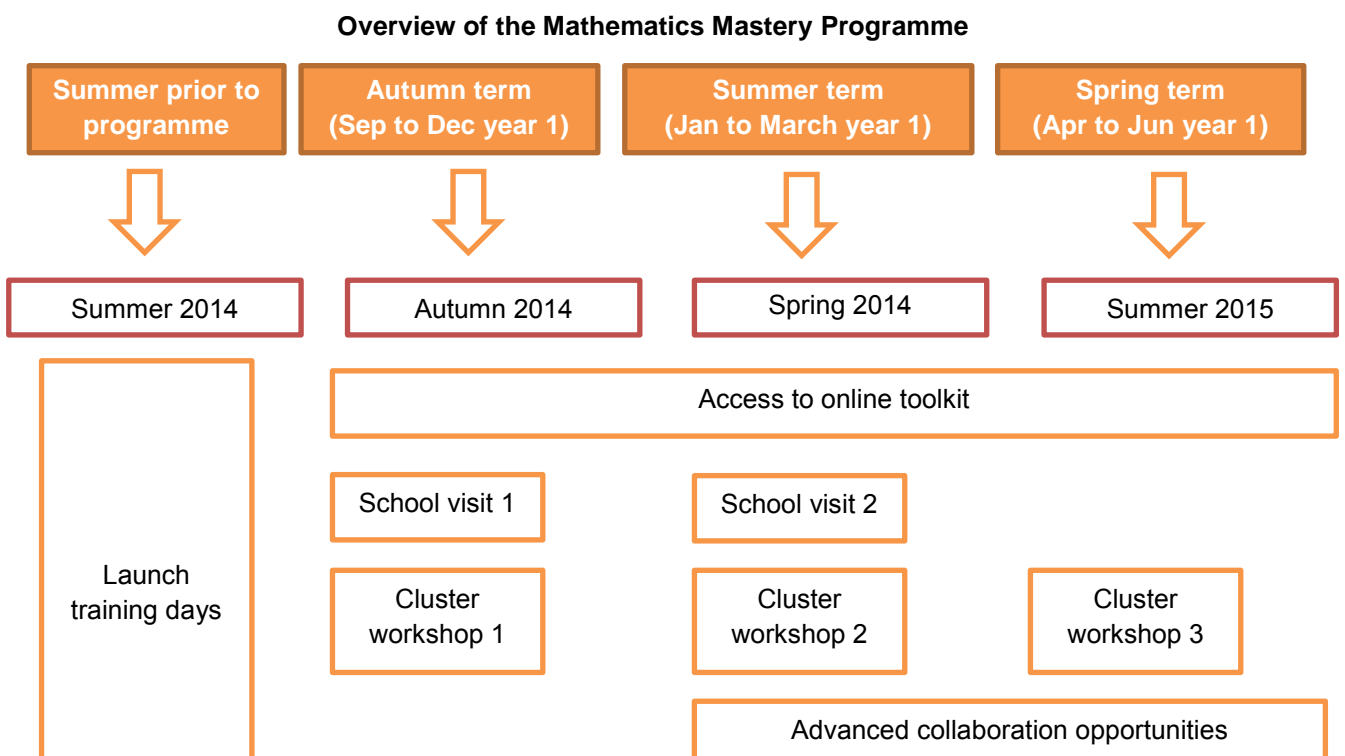
In addition to the 'mastery curriculum', other features of the approach include a systematic approach to mathematical language (see Hoyles, 1985; Lee, 1998), frequent use of objects and pictures to represent mathematical concepts (see Heddens, 1986; Sowell, 1989), and an emphasis on high expectations (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, so as to allow sufficient time for every child to access age-appropriate concepts and skills. Starting in Year 1, the main focus is to ensure all pupils have a firm understanding of number which will allow them to access and succeed in the other areas of maths. Schools roll out the approach to subsequent year groups year by year, with a view to transform maths achievement by Year 6.

Since taking part 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. In 2013, 95% of participating headteachers and school leads reported that the approach improved pupil attainment and the quality of teaching, and 98% said that it had increased pupils' enjoyment of maths.

Example problem

'Three consecutive numbers add up to 42. Find the numbers.'

Many pupils would tackle this using a trial and improvement approach (in later years perhaps eventually with algebra), but in a Mathematics Mastery classroom much emphasis is placed on different representations of numbers and the connections between them. Bar modelling is a key approach that is explicitly taught (as early as Year 2 for appropriately simple problems) and MM pupils may well use this approach:
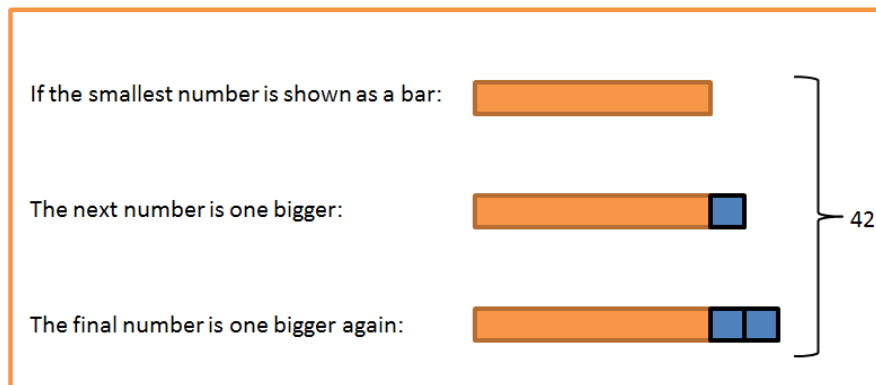
Pupils will see that if the total is 42, the total without the 'ones' is 39 and if the three bars add up to 39, each bar is worth 39 ÷ 3 = 13. So the numbers are 13, 14 and 15. This can easily be checked: 13 + 14 + 15 = 42.

The pupils' fluency in making connections stems from their learning from the beginning of the programme: when dealing with number bonds in the first term of Y1 they are taught using a 'part, part, whole' model to appreciate that as they know 7 = 5 + 2, they also know 5 = 7 – 2, 2 = 7 – 5 etc. Much of the number work covered in the early years of Mathematics Mastery is material expected to be covered in the next year, for example Y2 pupils comfortably tackling Y3 objectives.

The control condition was 'business as usual' during the intervention year (September 2012 to August 2013 for Cohort A and September 2013 to August 2014 for Cohort B). Control schools were then eligible to access the Mathematics Mastery programme after their year on the wait list had ended.

Note that no process evaluation was conducted for this trial and hence it is not possible to determine the extent to which individual schools adhered to the programme as it is described above.

Outcomes

The Number Knowledge Test (NKT) was used to examine children's skills in maths.[5] This is an individually administered oral test that takes about 10 minutes to complete, and requires only a few basic props (counters of two contrasting colours, visual displays with numbers). This test in particular was chosen as it was appropriate for reception age children and could be conducted orally. The test was administered by external testers from NatCen Social Research.[6] No specialist qualification is required to administer it—responses have only to be marked as correct or incorrect. A short training session was provided by the evaluation team to the test administration staff from NatCen Social Research to ensure that they understood and could effectively administer the test.

This test has been identified as highly predictive of achievement in primary maths (Cowan 2011). It derives from the theory that arithmetic skills are intimately related to the development of knowledge of the relations between numbers. In the Number Knowledge approach this is articulated as a central conceptual structure that is initially qualitative and subsequently becomes linear.

There are, however, limitations to the Number Knowledge Test (NKT). The test measures particular aspects of mathematical skill but not others. We do not know if the skills and knowledge measured by the test are necessarily the most important skills for subsequent general mathematics attainment. For example, the test does not assess some of the skills that underpin pre-algebra and these might better predict later mathematical attainment. Further, success in the test may well depend on other skills such as language skills, short term memory or executive function, or indeed self-efficacy. All of these factors also predict mathematics attainment which means that the test will potentially provide a summary measure of a range of skills acquired. Some caution therefore needs to be applied when interpreting the results of the trial and its implications for pupils' general mathematical attainment further on in their academic career.

An important consideration is whether the test is inherent to treatment—is it so closely related to the treatment as to be a questionable measure of the success of the trial? We have argued that the test is not specific to the Mathematics Mastery intervention in the sense that it was not developed for Mathematics Mastery, and moreover, the Number Knowledge Test was selected by an independent expert who had little riding upon the outcome of the study. However, it may be that the Mathematics Mastery curriculum is more closely related to the test than the curriculum followed by the control schools. For instance, Mathematics Mastery focuses particularly on knowledge of magnitude, place value and difference between numbers, and these aspects are measured by the test we use. The curriculum in the control schools, by contrast, may place less emphasis on these aspects of the curriculum and more emphasis on others, such as geometry, which are not tested by the NKT. This implies that it is possible that even if Mathematics Mastery had a positive impact on some aspects of pupils' mathematical skills as measured by the NKT, this may have been at the expense of the development of other mathematical skills not measured by NKT. The children in the control schools may have done better on these other aspects of the curriculum but they are not measured by the test we use. This implies again that we need to be mindful that the results may not be entirely generalisable to overall mathematics attainment.

Conversely, it should be noted that the test included a small number of concepts not included within the Mathematics Mastery curriculum during the trial period. It is possible that this could depress the effect picked up by the test. Further details about the test used are provided below.

---

[5] http://clArku.edu/numberworlds/nw_TestInfo.htm

[6] The baseline tests for Cohort A were conducted by Institute of Education MSc and PhD students. All other tests were administered by NatCen.

Test description and construct

A preliminary item asks the child to recite the count list from 1 to 10. Subsequent items are divided into four levels (0, 1, 2, 3) administered in that order:

- **Level 0** (approximate average level of a 4-year-old) items assess ability to count sets and subsets of items and judge relative magnitude.
- **Level 1** (approximate average level of a 6-year-old) items assess: basic addition and subtraction (for example 'How much is 2 + 4?'); knowledge of the number sequence ('What number comes two numbers after 7?'); relative magnitude of numbers ('Which is bigger: 7 or 9?'); and numerical distance ('Which number is closer to 7: 4 or 9?'). All items feature numbers less than 10.
- **Level 2** (approximate average level of an 8-year-old) items assess arithmetic, number sequence, relative magnitude, and numerical distance knowledge with two digit numbers.
- **Level 3** (approximate average level of a 10-year-old) assesses arithmetic and number sequence knowledge with more complex items. It includes items to assess the ability to compare differences (for example 'Which difference is smaller: the difference between 99 and 92 or the difference between 25 and 11?').

A discontinuation rule was applied: a child only received items from a later level if they demonstrated sufficient success in the earlier level. Points were awarded for correct answers. No timing was required and no strategy was forbidden, for example children could use their fingers to solve the arithmetic problems. Testers were guided as to how to stop the test appropriately if a child did not demonstrate sufficient success in a particular level.

The ability the test is supposed to assess is the central conceptual structure for whole number that is supposed to result from the integration of Number Knowledge (Case & Okamoto, 1996).

Test validity and reliability

A recent Economic and Social Research Council (ESRC) funded project (ESRC RES-062-23-0667, 'The development and importance of proficiency in basic calculation.') found that Number Knowledge predicted Year 4 teacher ratings of National Curriculum Levels for the three maths strands (correlations of between 0.68 and 0.70) and Year 4 standard scores on the WIAT Maths subscales (correlations of between 0.62 and 0.67). The interval of testing between Year 3 and 4 was one year.

Using item-response theory analysis, Gersten et al. (2007) found that the 'Number Knowledge' test has good reliability (0.94).

A study by Cowan (2011) in the UK found Cronbach alphas for 259 children of 0.82 (Year 3) and 0.81 (Year 4). Test–retest reliability was 0.61.

Floor effects

Ideally the tests we use to assess gains in mathematical knowledge should measure improvements irrespective of the child's initial level. Floor and ceiling effects can occur if the test is unable to discriminate between children with very low levels of skill or very high levels of skill (evidenced by a large number of children achieving the lowest score possible—floor effect, or the highest score—ceiling effect). The Number Knowledge test is designed for children from 4 to 8 years old. Children within the Mathematics Mastery trial were therefore in the middle of this range (Year 1, age 5–6). In the aforementioned ESRC project, no floor effects were observed, however there were ceiling effects for a small proportion of children in Year 3 (3.5%) and Year 4 (8.5%).

The Number Knowledge test has thus been chosen as it is a reliable measure of aspects of children's mathematics skills and is a strong predictor of children's later educational outcomes. Moreover, this test is not specific to the MM intervention.

Outcomes

- **Primary outcome:** total raw scores obtained across the whole of the Number Knowledge test.
- **Secondary outcome:** total raw scores obtained across the whole of the Number Knowledge test interacted with baseline Number Knowledge test score.

It is important to note that for this trial we were unable to ask for permission to link to the National Pupil Database. This means we do not have individual-level data to provide us with the characteristics of students, such as whether they were eligible for Free School Meals or their gender. Hence we are unable to undertake sub-analyses by gender or by FSM status.

Baseline test

Ideally, baseline measures (pre-tests) are taken prior to schools being randomly assigned to treatment or control. (This rules out the possibility that baseline test scores are influenced by the assignment to treatment or control groups.) However, given the short lead time into the project, this was not possible. Baseline tests were therefore conducted after random assignment occurred: they took place at the start of the school year, September to early October 2012 (Cohort A), and in September 2013 (Cohort B). Exactly the same test was used for the pre-test and the post-test (the 'Number Knowledge' test described above).

Sample size

Power calculations were based upon the following assumptions:

- A school level inter-cluster correlation (ICC) of 0.15 (i.e. $\rho = 0.15$).
- Equal cluster sizes of 30 pupils per school.
- 90 schools (clusters) participating in the trial.
- 50% of the variance in the post-test explained by the pre-test.
- 80% power for a 95% confidence interval.

Based upon these assumptions, the trial would be able to detect an effect of 0.18 standard deviations.

One can also estimate the ICC for the actual sample of schools/pupils that took part in the study. The ICC was estimated to be 0.11 for the post-test using a school fixed effect and 0.09 using a random effect (conditional upon the pre-test score). Moreover, our analysis confirms that 50% of the variance in the post-test could actually be explained by the pre-test. Cluster sizes did, however, turn out to be unequal, ranging from 21 to 116 pupils per school (the median school had 47 pupils). Using these figures, and if one were to estimate a random effects model, the actual minimum detectable effect was approximately 0.15.

Randomisation

All aspects of the randomisation were conducted by the independent evaluation team. A simple random sample was used, with allocation decided using a random number generator in Excel.

Analysis

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

$$Y_{ij}^{Post} = \alpha + \beta.Treat_j + \gamma.Y_{ij}^{Pre} + \varepsilon_{ij} \tag{1}$$

Where:

Y^{post} = child's post-test score on the Number Knowledge post-test;

Y^{pre} = child's baseline scores on the Number Knowledge pre-test;

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

ε = error term (with children clustered within school);

i = child I;

j = school j.

Post-test scores are now 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 = Hedge's G;

n_T = number of observations in the treatment group;

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

n_C = number of observations in the control group;

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

To account for the clustering of pupils within schools, the STATA survey (svy) command is 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 treatment?

The model presented above is estimated using Ordinary Least Squares (OLS) to investigate differences in mean test scores. We then investigate whether the Mathematics Mastery intervention had differential impact upon initial high and low achievers. We investigate this in two ways: (i) the

model presented in equation (1) is re-estimated using quantile regression and (ii) a second OLS regression model is estimated including a baseline test score – by – treatment interaction:

$$Y_{ij}^{Post} = \alpha + \beta.Treat_j + \gamma.Y_{ij}^{Pre} + \delta.Treat_j * Y_{ij}^{Pre} + \varepsilon_{ij} \qquad (2)$$

The coefficient of interest from equation (2) is δ—does the effect of the treatment depend upon the baseline achievement of the child?

Protocol

The protocol for this study is published online at:

http://educationendowmentfoundation.org.uk/library/maths-mastery-primary

Timeline

Cohort A
- Schools recruited from February to May 2012.
- Schools assigned to treatment or control group on 30 May 2012.
- Mathematics Mastery programme delivered between September 2012 and July 2013.
- Baseline tests conducted from September to October 2012.
- Post-tests conducted during July 2013.

Cohort B
- Schools recruited in April 2013.
- Schools assigned to treatment or control group on 16 April 2013.
- Mathematics Mastery programme delivered between September 2013 and July 2014.
- Baseline tests conducted during September 2013.
- Post-tests conducted during July 2014.

Impact evaluation

Participants

Sample allocation

Figure 1 provides details on sample allocation, attrition and potential contamination between treatment and control groups. A total of 90 schools were recruited to participate in the trial across the two cohorts. Schools were evenly split between treatment (n = 45) and control (n = 45) groups. All Year 1 children enrolled in one of the 90 participating schools and who completed the baseline test were considered to be part of the Mathematics Mastery trial. A total of 2,647 children were enrolled in the 45 treatment schools and 2,461 in the 45 control schools.

Attrition

Seven schools did not complete either the pre-test or the post-test (three treatment and four control).[7] A further two schools (one control and one treatment) completed the pre-test but not the post-test. The cluster (school) level response rate was therefore 92% (91% for the control group and 93% for the treatment group). This reduced the number of pupils to 2,386 in the treatment group and 2,244 in the control group.
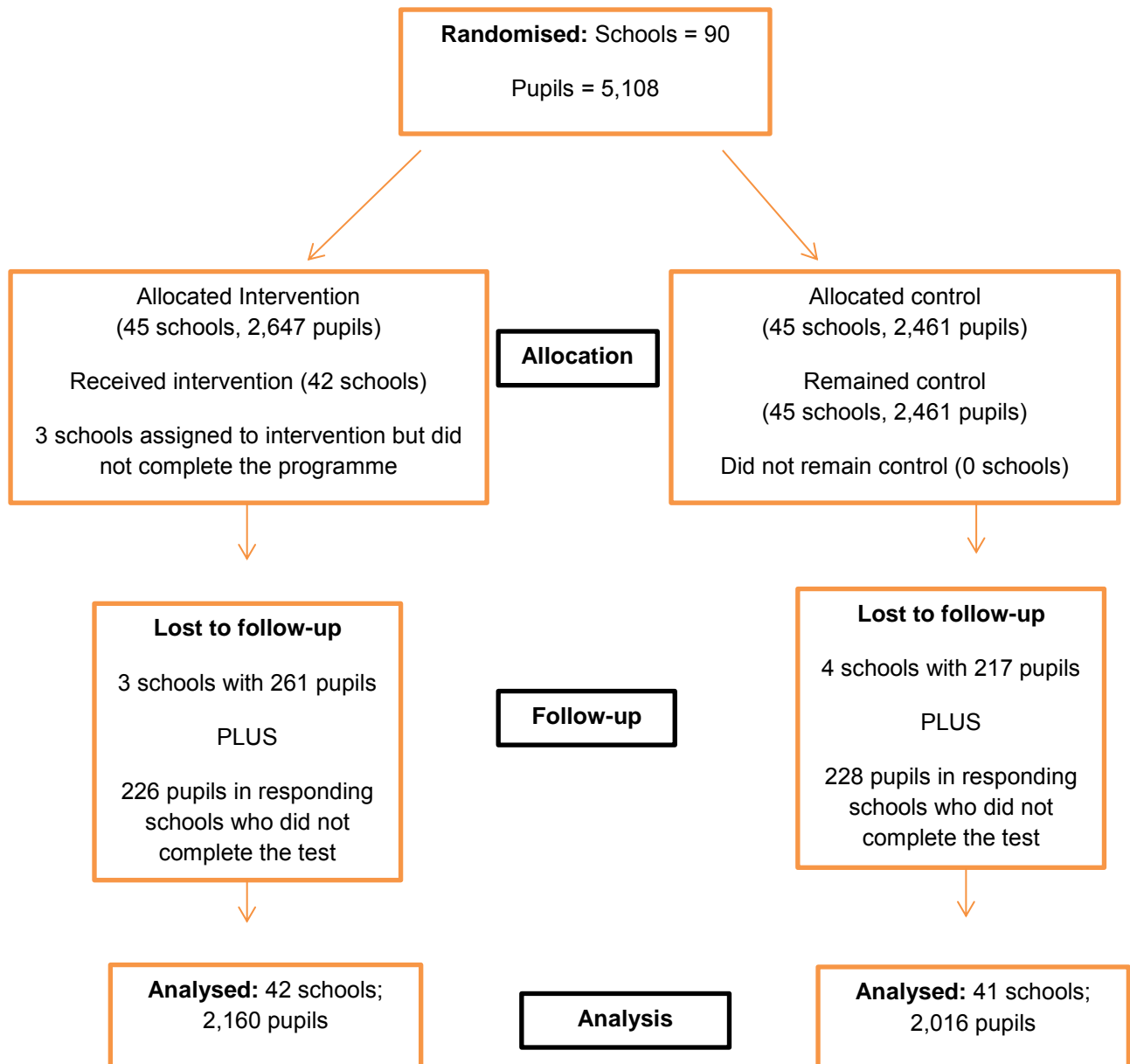
Moreover, within the 83 schools that conducted both the pre- and post-test, Number Knowledge scores were missing for some individual pupils. The most likely reasons were that these children were absent on the day of the post-test (e.g. through illness) or that they had moved to another school. The pupil response rate (within the 83 schools who conducted the post-test) was 90% in the control group and 91% in the treatment group. After excluding children with missing post-test data, pupil sample sizes fell to 2,160 in the treatment group (82% of the original allocation) and 2,016 in the control group (82% of the original allocation). These 4,176 children form the analysis sample.

Non-response analysis

Table 1 presents information on average baseline test scores for children that did not complete the post-test (either because their school withdrew from the study or because they were not in that school the day the post-test was conducted). These pupils were of notably lower ability than those children who did complete the post-test. Specifically, children who did not complete the post-test scored around a quarter of a standard deviations below the mean on the baseline test (-0.24 standard deviation for children in the control group and -0.27 for children in the treatment group). In contrast, those children who completed the post-test scored, on average, 0.03 standard deviations above the mean on the pre-test. A similar pattern was found in both Cohort A and Cohort B, though with there being slightly less evidence of selectivity in the former than the latter. Table 1 therefore suggests that attrition from the sample is not random—rather, lower-achieving children were more likely to have dropped out of the study than other groups.

---

[7] To get a proxy for the number of the pupils enrolled in each of these schools, information on the number of six year olds has been taken from EDUBASE - http://www.education.gov.uk/edubase/home.xhtml

Figure 1: Sample allocation and attrition

```
                    ┌─────────────────────────────┐
                    │  Randomised: Schools = 90   │
                    │                             │
                    │     Pupils = 5,108          │
                    └─────────────────────────────┘
```

Allocated Intervention (45 schools, 2,647 pupils) Received intervention (42 schools) 3 schools assigned to intervention but did not complete the programme	**Allocation**	Allocated control (45 schools, 2,461 pupils) Remained control (45 schools, 2,461 pupils) Did not remain control (0 schools)
Lost to follow-up 3 schools with 261 pupils PLUS 226 pupils in responding schools who did not complete the test	**Follow-up**	**Lost to follow-up** 4 schools with 217 pupils PLUS 228 pupils in responding schools who did not complete the test
Analysed: 42 schools; 2,160 pupils	**Analysis**	**Analysed:** 41 schools; 2,016 pupils

Table 1. A comparison of baseline achievement between children who did and who did not complete the post-test

	Respondent	Non-respondent (Control)	Non-respondent (Treatment)
Cohort A	**0.018**	**-0.159**	**0.000**
Cohort B	**0.039**	**-0.346**	**-0.403**
All pupils	**0.029**	**-0.242**	**-0.266**
Pupil n	**4,176**	**247**	**233**

Notes: *This table refers to pupils with valid baseline test data. It does not include children within the five schools that dropped out of the study before baseline testing took place. The 'non-respondent' group refers to children that completed the pre-test, but who did not complete the post-test. The sum of respondents and non-respondents does therefore not equal the total number of children initially enrolled in the trial. All figures reported in terms of effect sizes (standard deviation differences).*

Pupil characteristics

Table 2 compares children in the treatment and control groups in terms of their performance on the baseline test. This is done for the 4,176 children in the final 'analysis' sample (i.e. it includes only those children where both pre- and post-test scores are observed). There is evidence of imbalance between the treatment and control group. Table 2 illustrates that children in the control group score, on average, 0.07 below the sample mean. In contrast, children in treatment schools score, on average, 0.06 above the sample mean. Thus there is a difference in baseline test scores between treatment and control groups of approximately 0.13 standard deviations. (Although non-trivial in magnitude, this difference did not reach significance at the 10% threshold, t = 1.41; p = 0.16.)

Table 2. Comparison of baseline test scores between treatment and control groups

|	Control	Treatment	Difference
P10	**-1.35**	**-1.07**	**0.28**
P25	**-0.79**	**-0.50**	**0.28**
P50	**0.06**	**0.06**	**0.00**
Mean	**-0.07**	**0.06**	**0.13**
P75	**0.63**	**0.63**	**0.00**
P90	**1.20**	**1.20**	**0.00**
N	**2016**	**2160**	

Notes*: Authors' calculations. All test score data has been standardised to have a mean of 0 and standard deviation of 1 (across the 4,176 pupils in the analysis sample). Difference calculated as treatment minus control columns.*

Outcomes and analysis

Descriptive statistics

Figure 2 plots the distribution of post-test scores. Table 3 provides a further breakdown of the pre- and post-test score distribution for the treatment and control groups. There is little evidence of either floor or ceiling effects, with the distribution being broadly symmetric around the mean of 11 points for the pre-test and 15 test points for the post-test (standard deviation approximately 3.5 for the pre-test and 4 for the post-test). Figure 3 then plots the correlation between the pre-test and the post-test. Pre- and post-test performance is strongly associated, as one would expect, with the estimated Pearson correlation coefficient standing at 0.70 (Spearman's rank = 0.68). Finally, we estimate an OLS regression model with post-test scores as the dependent variable and the pre-test as the covariate. The R^2 is 0.496, suggesting that half the variance in the post-test can be explained by children's baseline test score.

Figure 2. The distribution of children's Number Knowledge post-test scores
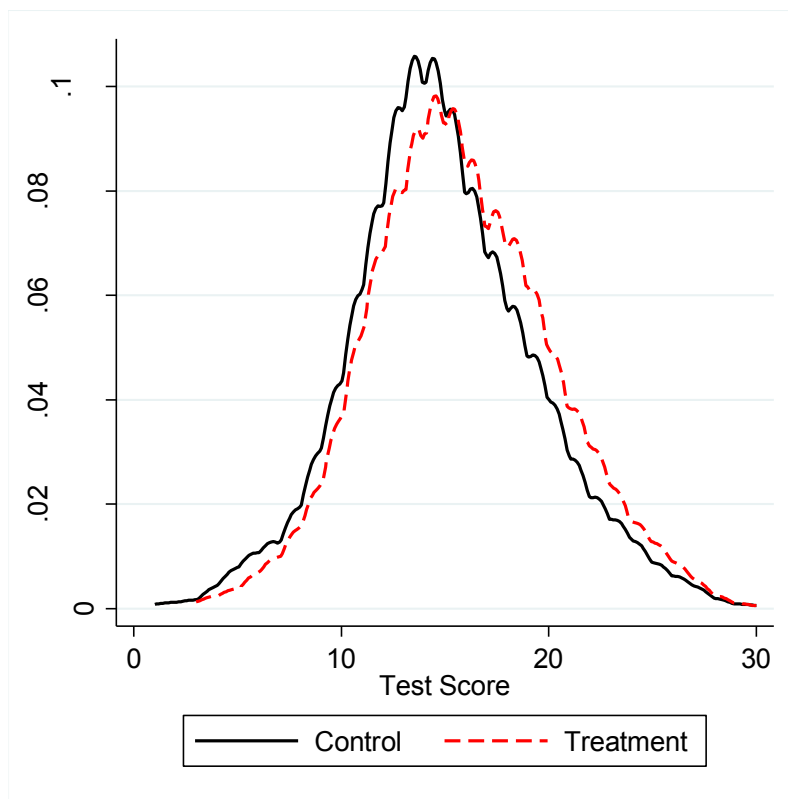
Notes: *Authors' calculations. Data restricted to analysis sample of 4,176 observations.*

Figure 3. Scatter plot between children's pre- and post-test score in the Number Knowledge test
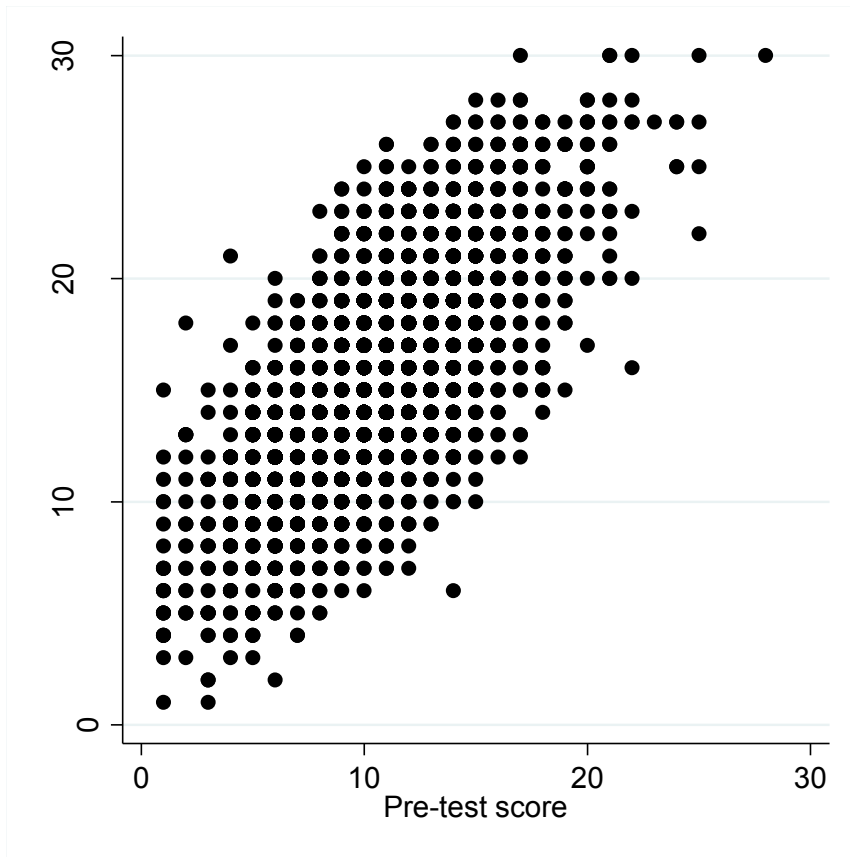
Notes: *Pearson correlation coefficient equals 0.70 and Spearman's rank 0.68.*

Table 3. Pre- and post-test score distribution for treatment and control groups
Notes: *Distribution of pre- and post-test scores for treatment and control groups. Figures refer to the 4,176 children included in the analysis sample.*

	Control				Treatment			
	Pre-test		Post-test		Pre-test		Post-test	
Mark	n	%	n	%	n	%	n	%
1	17	0.8	2	0.1	8	0.4	0	0.0
2	13	0.6	3	0.2	8	0.4	0	0.0
3	27	1.3	2	0.1	8	0.4	2	0.1
4	38	1.9	6	0.3	25	1.2	7	0.3
5	75	3.7	21	1.0	69	3.2	7	0.3
6	110	5.5	21	1.0	83	3.8	12	0.6
7	108	5.4	26	1.3	125	5.8	29	1.3
8	160	7.9	33	1.6	158	7.3	23	1.1
9	205	10.2	67	3.3	218	10.1	57	2.6
10	239	11.9	91	4.5	225	10.4	77	3.6
11	224	11.1	116	5.8	281	13.0	114	5.3
12	239	11.9	183	9.1	271	12.6	164	7.6
13	184	9.1	181	9.0	213	9.9	176	8.2
14	130	6.5	248	12.3	187	8.7	197	9.1
15	85	4.2	191	9.5	109	5.1	237	11.0
16	65	3.2	163	8.1	56	2.6	178	8.2
17	42	2.1	149	7.4	48	2.2	164	7.6
18	20	1.0	111	5.5	24	1.1	148	6.9
19	11	0.6	105	5.2	17	0.8	158	7.3
20	9	0.5	87	4.3	9	0.4	98	4.5
21	8	0.4	59	2.9	8	0.4	83	3.8
22	5	0.3	38	1.9	2	0.1	77	3.6
23	0	0.0	40	2.0	1	0.1	45	2.1
24	1	0.1	28	1.4	3	0.1	39	1.8
25	1	0.1	15	0.7	3	0.1	27	1.3
26	0	0.0	14	0.7	0	0.0	20	0.9
27	0	0.0	10	0.5	0	0.0	13	0.6
28	0	0.0	3	0.2	1	0.1	5	0.2
29	0	0.0	0	0.0	0	0.0	0	0.0
30	0	0.0	3	0.2	0	0.0	3	0.1
Total	2,016	100	2,016	100	2,160	100	2,160	100

Regression results: primary outcome

Parameter estimates from the OLS regression model are presented in Table 4. Results are presented separately for Cohort A, Cohort B and overall. The highlighted row shows the impact of the Mathematics Mastery intervention in terms of an effect size (Hedge's G).

Children who received the Mathematics Mastery intervention scored, on average, +0.10 standard deviations higher on the post-test. This, however, only reached statistical significance at the 10% level (t = 1.82; p = 0.07), with the 95% confidence interval ranging from -0.01 to +0.21. Within Cohort A, children in the treatment group scored (on average) +0.09 standard deviations above those children in the control group (confidence interval -0.06 to +0.24). The analogous effect in Cohort B was +0.10 (confidence interval -0.05 to 0.26). Consequently, although the Mathematics Mastery intervention may have had a small positive effect on children's test scores, it is not possible to rule out sampling variation as an explanation.

Table 4. Regression results

	Cohort A		Cohort B		Overall	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**
Intervention Group (Ref: Control)						
Treatment	**0.091**	**0.075**	**0.105**	**0.078**	**0.099***	**0.054**
Pre-test score	**0.695****	**0.024**	**0.711****	**0.025**	**0.704****	**0.016**
Constant	**-0.055**	**0.052**	**-0.048**	**0.050**	**-0.051**	**0.036**
N	**1,868**		**2,308**		**4,176**	

Notes: *Authors' calculations. Dependent variable is total Number Knowledge raw score (standardised to mean 0 and standard deviation 1). SE stands for standard error. * and ** indicate statistical significance at the 10% and 5% levels, respectively. Standard errors clustered at the school level.*

Secondary outcomes

Table 5 illustrates regression results when a treatment*pre-test score interaction is included in the model. This will illustrate whether the treatment had differential effects for children of different baseline ability. Estimates are again presented in terms of effect sizes. The interaction effect is not significantly different from zero, with the 95% confidence interval ranging from -0.01 to +0.02. Thus there is little evidence that the effect of Mathematics Mastery differs between children with different levels of prior achievement.

Table 5. Regression results: treatment * pre-test interaction

	Cohort A		Cohort B		Overall	
	Beta	SE	Beta	SE	Beta	SE
Intervention Group (Ref: Control)						
Treatment	0.217	0.142	-0.059	0.135	0.069	0.098
Pre-test score	0.715**	0.033	0.685**	0.028	0.699**	0.022
Pre-test * Treatment interaction	-0.012	0.014	0.015	0.011	0.003	0.009
Constant	-0.053	0.053	-0.050	0.050	-0.051	0.036
N	1,868		2308		4,176	

Notes: *Authors' calculations. The dependent variable is children's score on Number Knowledge post-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 at the school level.*

In exploratory analysis,[8] Figure 4 examines whether there is any evidence of heterogeneity of the treatment effect across the 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. Recall that the Mathematics Mastery intervention is particularly concerned with the 'mastery' of basic skills, 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. There is some, but relatively little, evidence that the intervention was less effective for the bottom half of the test distribution. The quantile regression estimates all sit relatively closely to the dashed horizontal line (OLS estimate), with all those below p40 or above p65 not statistically different from zero at even the 10% level. Overall, there is little evidence of heterogeneous treatment effects occurring across the post-test distribution.

---

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

Figure 4. Quantile regression estimates of the effect of the Mathematics Mastery intervention across the Number Knowledge test score distribution
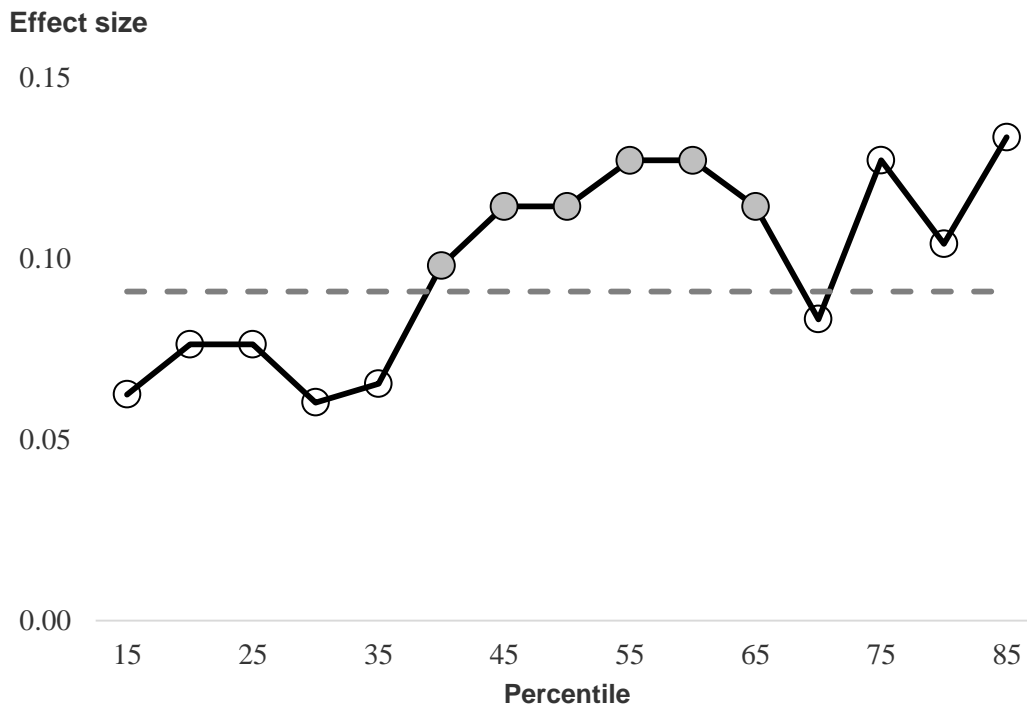
Notes: *Authors' calculations. Dashed horizontal line illustrates the OLS estimate. Solid grey circular markers indicate whether the treatment effect is significantly greater than 0 at the 10% level. Standard errors bootstrapped by cluster using 100 replications. Dependent variable is children's score on the Number Knowledge test. Open circles represent coefficients that are not statistically significantly different from 0 at the 10% significance level.*

Cost

The costs reported below come from information provided by Ark and have not been audited by the evaluation team.

For two-form primary schools, there is an upfront cost of £6,000 for participating in the programme. (This is an 'at cost' price charged by Ark to cover basic infrastructure.) A total of seven days staff time is required for training: one day for the headteacher, two days for the head of maths, and two days for two maths teachers. To calculate the cost of headteachers' time, we take the 20th percentile of the headteacher pay scale in England and Wales (£56,027):[9] this is then divided by 230 (the approximate number of working days in a year) to give a headteacher day rate of £243.[10] We then inflate this figure by a fifth to allow for other costs not directly incorporated into headteachers' salaries (such as employer contributions to pensions) giving a total cost of £292.[11] Analogous calculations were made for the head of maths (two days training at a final day rate of £196)[12] and the class teachers (a total of four days training at a final day rate of £141).[13] Total training costs therefore amount to £1,248. The

---

[9] This information has been drawn from http://www.education.gov.uk/get-into-teaching/about-teaching/salary/pay-and-benefits

[10] The headteacher pay scale in England and Wales (outside of London) ranges from £107,210 to £43,232. Primary head teachers are on average at the lower end of this scale and we have assumed on average at the 20th percentile. We have assumed headteachers work 46 five day weeks per year (with the other six weeks as holiday).

[11] 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.

[12] We have assumed this teacher to be an experienced senior teacher at the top of the upper pay scale (£37,496).

[13] It is assumed the teacher's will be on the 'main' pay scale, which ranges from £22,023 to £32,187 (median 27,105).

total cost to the primary school is £7,248 for this first year. We then estimate the average number of pupils affected per primary school as 57—the number of pupils initially enrolled into the primary trial (5,108) divided the number of primary schools initially enrolled (90). The 'per pupil' cost of delivering the primary school intervention was therefore £127 for the first year. In subsequent years schools are able to opt for different pathways depending on the amount of support and training they wish to choose; they also have ongoing access to the curriculum materials for additional year groups. The per pupil cost therefore reduces considerably, to below £30 per pupil for additional year groups.

Conclusion

Key conclusions
1. On average, Year 1 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 little evidence that the effect of Mathematics Mastery differs between children with different levels of prior achievement.
3. Combining the findings from this study and a second randomised controlled trial of Mathematics Mastery involving Year 7 pupils strengthens the overall evidence for the approach.
4. Given the low per-pupil cost, Mathematics Mastery may represent a cost-effective change for primary schools to consider.
5. It would be worthwhile to track the medium and long-term impact of the approach, in part to assess the degree to which the test used in this evaluation is predictive of general mathematics attainment and 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 stand out:

Short-term follow-up
Pupils' progress was evaluated after just one academic year. This may be considered a relatively small 'dose' of the Mathematics Mastery programme. Little is known about the possible *cumulative* impact of the programme on children's maths ability over a prolonged period of time.

A new programme in schools
The Mathematics Mastery intervention introduced a new approach to teaching maths within treatment schools. Teachers within these schools had to get used to the Mathematics Mastery curriculum and learn how to teach it effectively. In contrast, teachers in control schools proceeded with 'business as usual'—teaching the same material as they had done in previous years. With more experience teaching the programme it is possible that teachers would become more effective in implementing it.

Test validity
The test used has been found to predict future mathematical attainment. however the test does not measure all aspects of pupils' mathematics skills. It is possible that pupils in control schools following a broader curriculum developed better mathematics skills in aspects of the curriculum not covered by the Mathematics Mastery intervention schools. The test may favour Mathematics Mastery schools by being more closely related to the topics being taught in the intervention. However, it should also be noted that some test questions included material not covered within the Mathematics Mastery curriculum during the trial period.

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.

Balance

There was some evidence of imbalance between treatment and control children in terms of their baseline test scores (approximately 0.13 standard deviations). Although this was controlled for in the analysis, it was not possible to take into account other possible sources of imbalance such as ethnicity or Free School Meal status. This was due to this being one of the first EEF trials commissioned, where the processes for linking to National Pupil Database (NPD) data had not been agreed. (Consequently, it has not been possible to link data from this trial to the NPD, and thus characteristics such as free school meal eligibility and ethnicity remain unobserved.)

Recommendations

More generally, subsequent trials could undertake a process evaluation so that fidelity analysis can be conducted. This will help to determine how closely teachers adhere to the intervention design and how the pedagogy and curriculum covered differs from the control schools. Further, to ensure the robustness of the findings, future work might use different outcome measures that assess different aspects of pupils' mathematical skills. Only then will we be sure that the results here are not merely the result of a specific test being used which relates to the curriculum covered by Mathematics Mastery, but rather the result of a better pedagogical approach.

Interpretation

The hypothesis tested in this study was that the first year of the Mathematics Mastery programme would lead to a significant improvement in Year 1 (age 5 to 6) children's maths test scores. Although point estimates were consistent with a small, positive gain, the study only had sufficient statistical power to rule out sampling variation as an explanation at the 10% significance level. Within the context of the wider educational literature, the effect size reported (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). This difference may arise for a number of reasons. For example, the intervention here may have lacked some of the effective Mathematics Mastery elements that these other trials included. Without further research it is not possible to explain why the results from this trial are more modest than this previous research.

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 an older cohort of children (age 11 to 12), a similar effect size has been found (approximately 0.06 standard deviations for a one-year dose). 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 the 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.[14]

Third, children are likely to follow the Mathematics Mastery programme for a number of years (perhaps throughout primary school), whereas this evaluation has considered the impact of just the first year of the programme. Long-run effects after sustained exposure to the programme could be significantly higher, and will be assessed in a follow-up study using Key Stage 2 data.

---

[14] Although an entirely different treatment, 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). The Literacy Hour and the MM interventions do differ but this is an interesting point to note.

Future research and publications

There are three key questions for future research to address. First, this report has focused on the impact of the first year of the Mathematics Mastery intervention, with outcomes measured immediately after the RCT had 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.

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 7. 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.

Case, R. and Okamoto, T. (1996) 'The Role of Central Conceptual Structures in the Development of Children's Thought'. *Monographs of the Society for Research in Child Development,* 61 (1/2): 1–295.

Cowan, R. (2011) 'The Development and Importance of Proficiency in Basic Calculation'. Accessed 21 / 08 / 2014 from www.ioe.ac.uk/Study_Departments/PHD_dev_basic_calculation.pdf

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

Freudenthal, H. (1968) 'Why to Teach Mathematics so as to Be Useful'. *Educational Studies in Mathematics,* 1, 3–8.

Gersten, R., Clarke B. and Jordan N. (2007) 'Screening for Mathematics Difficulties in K–3 Students'. Centre on Instruction Research Report, accessed 05/12/ 2014 from http://files.eric.ed.gov/fulltext/ED521575.pdf

Guskey, T. and Pigott, T. (1988) 'Research on Group-Based Mastery Learning Programs: A Meta-Analysis'. *The 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): 205–14.

Jerrim, J. and Vignoles, A. (2015—forthcoming) 'The Causal Effect of East Asian "Mastery" Teaching Methods on English Children's Mathematics Skills'. DoQSS working paper series.

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' in *Studies in Educational Leadership*, Springer.

OECD (2013) 'Programme for International Student Assessment 2012: Results from PISA 2012. United Kingdom Country Note'. Downloaded 16/11/2014 from www.oecd.org/unitedkingdom/PISA-2012-results-UK.pdf

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

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

Wise, S. and DeMars, C. (2005) 'Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions'. *Educational Assessment*l, 10(1):1–17.

Waxman, H., Wang M., Anderson K. and Walberg H. (1985) 'Adaptive Education and Student Outcomes: A Quantitative Synthesis'. *The Journal of Educational Research* 8(4): 228–236.

Appendix A: Padlock rating
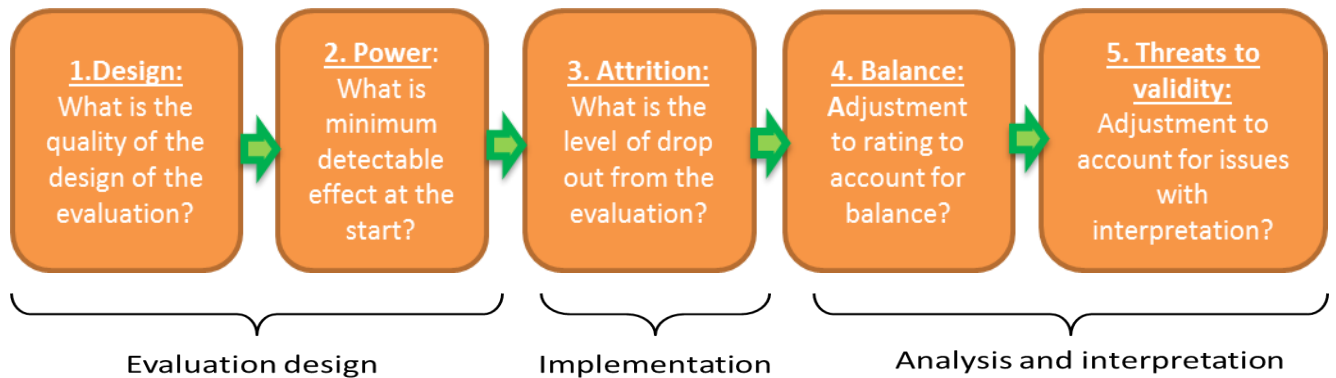
Rating	1. Design	2. Power (MDES)	3. Attrition	4. Balance	5. Threats to validity
5 🔒	Fair and clear experimental design (RCT)	< 0.2	< 10%	Well-balanced on observables	No threats to validity
4 🔒	Fair and clear experimental design (RCT, RDD)	< 0.3	< 20%		
3 🔒	Well-matched comparison (quasi-experiment)	< 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 3 🔒. This means that the conclusions have moderate security.

The trial was designed as an efficacy trial and could achieve a maximum of 5 🔒. This was a large and well conducted randomised controlled trial. There was little attrition (7 of 90 schools dropped out – 8%. 4 of the 7 schools that dropped out were in the control arm). There was higher attrition at the pupil-level (18%). There was considerable imbalance in the baseline test score between treatment and control in the pupils that were analysed (i.e. those that also had a post-test). Pupils in the treatment arm had on average 0.13 standard deviations higher pre-test scores than in the control group, which reduced the number of padlocks by two. There were a few small threats to validity, such as potential for treatment inherent measure and pre-testing after randomisation. Therefore, the overall padlock rating is 3 🔒.

Appendix B: 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 C: 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).

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 five per cent 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 A1. Meta-analysis results

	Primary school	Secondary school	Meta (combined)
Number of schools	83	44	127
School response rate	92 per cent	88 per cent	-
Number of pupils	4,176	5,938	10,114
Pupil response rate	82 per cent	77 per cent	-
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	0.004 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 per cent and five per cent levels respectively.

The Education Endowment Foundation

9th Floor, Millbank Tower

21–24 Millbank

London

SW1P 4QP

www.educationendowmentfoundation.org.uk