The link between East Asian 'mastery' teaching methods and English children's mathematics skills

John Jerrim[1]

Anna Vignoles[2]

1 Institute of Education, University of London

2 University of Cambridge

October 2015

http://johnjerrim.com/papers/

Abstract

A small group of high-performing East Asian economies dominate the top of the Programme for International Student Assessment (PISA) rankings. Although there are many possible explanations for this, East Asian teaching methods and curriculum design are two factors to have particularly caught policymakers' attention. Yet there is currently little evidence as to whether any particular East Asian teaching method actually represents an improvement over the status quo in England, and whether such methods can be successfully introduced into Western education systems. This paper provides new evidence on this issue by presenting results from two clustered Randomised Controlled Trials (RCT's), where a Singaporean inspired 'mastery' approach to teaching mathematics was introduced into a selection of England's primary and secondary schools. We find evidence of a modest, positive treatment effect that comes at a relatively low per-pupil cost.

Key Words: Maths Mastery; Randomised Controlled Trial; Singapore; PISA

Contact Details: John Jerrim (J.Jerrim@ioe.ac.uk) Department of Quantitative Social Science, Institute of Education, University of London, 20 Bedford Way London, WC1H 0AL

Acknowledgements: This paper includes analysis from two evaluations of the Mathematics Mastery programme funded by the Education Endowment Foundation. The cost-benefit analysis included in this paper was not funded by the Education Endowment Foundation. Professor Richard Cowen, Institute of Education, provided advice on the appropriate mathematics tests to use for the RCT. Meg Wiggins, Anne Ingold and Helen Austerbury assisted with the secondary school testing. Valuable comments have been received from John Micklewright and Emla Fitzsimons. The intervention was delivered by ARK schools http://www.arkschools.org/. Funding for the evaluation was provided by the Education Endowment Foundation.

1. Introduction

The Programme for International Student Assessment (PISA) is a major cross-national study of school children's academic achievement. Since its inception in 2000, its ranking of the world's education systems has drawn the attention of academics, educationalists, journalists and policymakers alike. A small group of high-performing East Asian economies (e.g. Singapore, Japan, Hong Kong, South Korea) consistently dominate the top of these international 'league tables'. This is particularly true in mathematics, where children from such countries are, on average, more than one school year ahead of their Western peers. Consequently, two of the most frequently asked questions by education policymakers today are '*what drives East Asian educational success'* and '*what can we do to catch up*'?

<< Table 1 >>

There are, of course, several explanations as to why PISA test scores differ between countries in the East and the West. This point is illustrated in Table 1, which compares various aspects of the education systems in England and Singapore (two countries of particular interest within this paper). A number of substantial differences exist, including school resources, provision of out-of-school tuition and school discipline. Yet, despite these many differences, it is teaching methods and design of the curriculum that has particularly caught policymakers' attention. For instance, to inform upcoming changes to the mathematics curriculum in England, the Department for Education (2012) conducted an extensive review of the mathematics syllabus in a number of East Asian countries. Similarly, a selection of British officials have visited East Asian economies to observe their teaching practices (Department for Education 2014a), under the presumption that this is driving their educational success. Indeed, as Liz Truss (former Under Secretary of State for Education in England) noted of one such visit:

*'this represents a real opportunity for us to see at first hand the **teaching methods** that have enabled their young people to achieve so well in maths.*' [Emphasis our own].

As a consequence of such visits, the Department for Education has now set-up an exchange programme, where teachers from East Asia are being flown into England to demonstrate and apply their teaching methods within this country's schools (see Department for Education 2014b). Thus, despite difficulties in even defining the concept of an 'East Asian teaching

method', policymakers continue to believe this to be a key reason why mathematics achievement is so much greater in the East than the West.

Yet simplistic attempts to 'borrow policy' from other countries is problematic (Crossley and Watson 2009). Two particular issues stand out. The first is causality. There are significant cultural, economic and historic differences between countries, as well as a number of differences in how the education system is designed and managed (see Table 1). It is therefore almost impossible to tell from studies like PISA what is leading to the cross-national variation in children's test scores. Indeed, there is very little evidence that East Asian teaching methods, however defined, are actually superior to those currently being used in England's (or other Western countries) schools. Second, even if some East Asian teaching methods are potentially more effective than those used in England, one simply does not know whether they can be successfully implemented within the English, or, indeed, other, educational system.

This paper does not therefore attempt to determine whether so called 'East Asian teaching methods' can improve children's achievement in England, especially since it is not even clear what these methods are. Instead this paper provides (to our knowledge) the first evaluation of how introducing a specific teaching approach, inspired by current practise in Singapore, influences achievement within England's schools. This is done via estimation of the causal effect of the 'Maths Mastery' teaching programme after it had been implemented within a selection of England's primary and secondary schools for one academic year. This particular programme is based upon approaches to teaching mathematics in Singapore (ranked 2nd out of 65 economies in the PISA 2012 mathematics rankings) and, potentially at least, represents a radical change to standard practise in England (see Guskey 2010). In particular, fewer topics are covered in greater depth, with every child expected to reach a certain level (i.e. to 'master the curriculum') before the class progresses on to the next part of the syllabus together. The notion that Singaporean teachers place more emphasis on whole class mastery of concepts is indeed supported by the Teaching and Learning International Survey (Micklewright et al 2014). This survey indicates that, whereas three-in-five teachers in England differentiate their lessons for pupils with different abilities, only one in five Singaporean teachers do (also see Table 1). Greater emphasis is also placed upon children's problem solving skills, with this complemented by an integrated professional development programme for teachers, and the sharing of best practise amongst a network of schools.

This paper reports results from two field experiments designed to estimate the effect of a one year exposure to this programme. A clustered Randomised Controlled Trial (RCT) methodology is used, involving more than 10,000 pupils enrolled in 90 English primary schools and 50 secondary schools during the 2012/13 and 2013/14 academic years. Both the primary and the secondary school trials suggest a positive impact of the programme, though with significant heterogeneity by 'school quality', particularly within the latter. It is perhaps noteworthy that the magnitude of the effect found is similar to that for some other curriculum and pedagogical interventions also attempting to improve basic skills. This includes the 'The Literacy Hour' - a change made to the English curriculum made in the late 1990's – which was found to have a small, positive impact by Machin and McNally (2008).

The paper proceeds as follows. Section 2 overviews the Maths Mastery (MM) intervention, with our empirical methodology detailed in section 3. Section 4 provides results from the two RCTs. Conclusions and directions for future research follow in section 5.

2. The Maths Mastery intervention

Maths Mastery is delivered in England by the academy chain ARK. Our study considers two particular versions of their programme; one appropriate for Year 1 pupils (age 5/6) and one appropriate for year 7 pupils (age 11/12). The introduction of a 'mastery curriculum' is central to the Maths Mastery approach. This is where the vast majority of pupils' progress through the curriculum at the same pace, with subject matter and learning content broken into units with clearly defined goals. Academically weaker pupils are expected to reach a basic standard in each unit before the whole class moves on to the next topic together. In the meantime, more able pupils are encouraged to explore the current learning unit in more depth. (This is in contrast to standard practise in England, where more able pupils are accelerated on to learning a new topic). It is thought that this approach reduces the need to repeatedly revisit material, and promotes depth of understanding over memorised procedures.

Other features of MM include a systematic approach to mathematical language (Hoyles 1985), frequent use of objects and pictures to represent mathematical concepts (Heddens 1986; Sowell 1989), and an emphasis on high expectations and a 'growth' mind-set (Dweck 2006; Boaler 2010). For younger children, this translates into prominent use of objects and pictures to illustrate numbers. The approach also prioritises problem solving skills, and

encourages deep understanding of mathematics over procedural knowledge (Skemp 2006). Every pupil is therefore expected to understand what they are doing, rather than just learning to repeat routines. This in turn means they are better equipped to apply this knowledge when solving numerical problems.

Figure 1 provides an example of the Maths Mastery approach. In this, children are asked:

'*There are three consecutive numbers that add up to 42. What are these numbers?*'

Young children in England would typically use an iterative 'trial and improvement' method to answer this question. In contrast, Maths Mastery emphasizes the representation of numbers and connections between them. A prime example of this is 'bar-modelling' as illustrated in Figure 1. (This technique is typically taught to children from around age 6 or 7). Maths Mastery pupils will recognise that, as the total is 42, the total without the 'ones' (i.e. the grey portions of the bars) equals 39. Then, if the three sections add up to 39, each must be worth 13 (since 39 ÷ 3 = 13). As the question states the numbers are sequential, children then immediately reach the answer of 13, 14 and 15. Further details and examples can be found at www.mathematicsmastery.org/.

<< Figure 1 >>

In the short-term, it is challenging for schools and teachers to move to such a different approach. Therefore, to assist the transition, ARK offers schools the following support during the first 'moving to mastery' year:

- **Training and in-school support**. Before the programme begins, school leaders, maths coordinators and class teachers receive either one or two days of training. This is followed by two in-school development visits, three multi-school cluster workshops, and access to an online toolkit. (This includes detailed information on continuous professional development resources, assessments and leadership frameworks).
- **Curriculum-embedded continuous professional development**. Teachers are supported to put the principles into practice through 'lesson designs.' These adapt to the needs of each class via the online toolkit, in the hope that teachers can use lesson planning time to also develop as professionals.

- **Collaboration and peer support**. Teachers from different schools are encouraged to collaborate to develop best practice (Mulford, Silins and Leithwood 2004). This is via both face-to-face and online interaction, with a focus upon sharing ideas and supporting one another in applying the approach.

Within our two RCT's, schools in receipt of the Maths Mastery programme had access to this support. A timeline of when activities were provided to schools can be found in Figure 2.

<< Figure 2 >>

It is not possible to specify precisely the extent to which the Maths Mastery intervention described here approximates the teaching practices and curricula used in East Asian countries such as Singapore (the country from which this intervention draws most heavily). This is because there is no one pedagogy or curricula used across schools in East Asia, though many teachers do use the principles adopted in MM. Clearly the English context is also very different and factors such as teacher familiarity with the basic ideas, teaching quality more generally, and other classroom practices, will differ considerably between England and countries such as Singapore. This will influence the way in which the programme is applied in England. Further some specific Singapore practices, such as the use of very extensive homework, have not been adopted to the same extent in the sample of English schools, an issue we return to below. Perhaps above all one must remember that the Singapore approach includes more 'drill style' pedagogy than the average teacher in England will be accustomed to. Moreover, many will have explicitly rejected such approaches during their teacher career, again an issue we discuss later. Given all this, one cannot claim that this evaluation assesses the potential impact of East Asian or Singaporean pedagogy in English schools per se. Rather, we evaluate the impact of a programme modelled on, but not synonymous with, Singaporean mathematics teaching approaches.

It is also important to understand that Maths Mastery is designed to take a long term view of transforming maths achievement. The curriculum is cumulative, thus allowing every child sufficient time 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. This then allows them to access and succeed in the other areas of mathematics. Schools roll out the approach to subsequent year groups, with a view to transforming achievement by the end of Year 6 (i.e. five years after the programme was first introduced in schools). It has not been possible to evaluate the cumulative effect of Maths Mastery over five years using an experimental

design. Our evaluation therefore considers the impact of a relatively small (one year) dose of the MM intervention, immediately at the end of the first year.

3. Data and Methods

Overview

In this paper we report results from two clustered Randomised Controlled Trial (RCT) of the Maths Mastery programme in England. One RCT was conducted in the first year of primary school (5/6 year olds) and the other in the first year of secondary school (11/12 year olds) [1].

The primary trial involved two school cohorts. Cohort A consisted of 40 schools conducting the trial in the 2012/13 academic year, with a further 50 schools participating in 2013/14 (Cohort B). Half of these 90 schools were randomly assigned to receive the Maths Mastery programme, with the other 45 schools assigned as controls. The secondary school trial ran during the 2013/14 academic year, and involved 25 treatment schools and 25 control schools.

All pupils within treatment schools were taught using the Maths Mastery approach. Control schools were asked to proceed with 'business as usual', meaning they would continue to use the same curriculum and approaches to teaching mathematics that they had used in previous years. An accompanying qualitative evaluation of the trial found that 'business was indeed as usual in control schools', with little evidence of any change in practise before and after the intervention took place (see Jerrim et al 2015).

Recruitment

Recruitment of schools into the trial was undertaken by the charity responsible for delivering the intervention (ARK). Schools were not randomly selected to take part; rather, they were purposefully recruited. Only two formal exclusion criteria were set. First, the schools could not already be using the Maths Mastery programme. Second, schools could not be from the private sector. Otherwise, the charity was free to recruit any school within England. However, the organisation funding the trial (the Education Endowment Foundation), made clear their preference for the schools selected to have a high proportion of children from low socio-economic backgrounds.

The main implication of this recruitment process is that the 'external validity' of the two trials will be relatively low. In other words, we will only be able to estimate the impact of

[1] The protocol (pre-analysis plan) for this study is published online at: http://educationendowmentfoundation.org.uk/library/maths-mastery-primary

Maths Mastery amongst schools who were willing to participate in the trial, with our findings not necessarily generalizable to the wider population.

Nevertheless, we are able to provide further information on the characteristics of children and schools who took part in the RCTs, and how this compares to the state school population for England. Details are provided in Table 2. Panel A provides evidence for the 90 schools recruited to the primary school trial[2]. The left hand side refers to cohort A and the right hand side to cohort B. The gender and month of birth distributions for children enrolled in the trial is very close to that for the population as a whole. There are, however, a greater proportion of children eligible for Free School Meals (a marker of low income) enrolled in the trial than in the national population. The final six rows refer to children's scores on the Foundation Stage Profile – six teacher-based assessments of children's development at approximately age 5. (We have standardised each of these scales to mean 0 and standard deviation 1 across the population). Interestingly, whereas cohort A children scored below the national mean on each of these scales (typically by around 0.10 standard deviations), cohort B children tend to score above the mean (again by around 0.10 standard deviations). Hence we cannot say that the schools recruited are typical of schools in England. Rather, they are somewhat more disadvantaged, while also potentially differing from the wider population in certain unobserved ways.

<< Table 2 >>

Table 2 Panel B provides a similar analysis for the 50 secondary schools initially recruited. A total of 29 per cent of children enrolled in the trial were eligible for Free School Meals, compared to 18 per cent 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 per cent in the sample versus 5 per cent in the population) and Asian (20 per cent in the sample versus 10 per cent in the population) groups. Trial participants also tended to have lower Key Stage 1 (age 7) and Key Stage 2 (age 11) scores than the state school population as a whole[3]. For instance, their Key Stage 1 average points scores (and Key Stage 2 maths test scores) were approximately 0.2 standard

[2] It has not been possible to link administrative records to individual pupils within the primary school trial. The figures presented are therefore based upon administrative records held by the school, based upon the autumn census enrolment data. Figures on pupil enrolment therefore differ slightly to those provided for the primary school trial sample provided in the main text.

[3] Key Stage 1 and Key Stage 2 exams are national tests all state school children in England sit at age 7 and age 11.

deviations (0.1 standard deviations) below the population mean. This seems to be driven, at least in part, by the fact the trial particularly under-represented high achievers (relative to the population). For instance, just 12 per cent of children participating in the trial were award Level 3 in their Key Stage 1 maths test, compared to 19 per cent of all state school pupils in England. The sample of secondary schools recruited therefore appears more disadvantaged than the national average.

Attrition

A total of 90 primary schools containing 5,108 pupils (2,647 treatment and 2,461 control) were initially recruited into the primary school trial. Seven of these schools (three treatment and four control) dropped out of the study. Moreover, a small number of children in each school did not complete the post-test, due to either absence on the day of the test (e.g. through sickness) or having moved to another school. The final response rate was therefore 92 percent at the school level and 82 percent at the pupil level. Our final analysis includes the 1,868 pupils in the treatment group and 2,308 in the control group for whom both pre and post-test scores were available. See Appendix A for further details.

Analogous figures for the secondary school trial were 50 schools initially recruited containing 7,712 children (4,004 treatment and 3,708 control). Six schools dropped out of the study, with final response rates of 88 percent at the school level and 77 percent at the pupil level. The final pupil sample size was 5,938 pupils (3,251 treatment and 2,687 control) within 44 schools (23 treatment and 21 control). See Appendix B for further details.

Testing

The tests used in both the primary and secondary trials were selected by the evaluation team, independently of ARK who developed and delivered the Maths Mastery programme.

In the primary school trial, all children were tested at the start and end of the academic year using the 'Number Knowledge' test (Okamoto and Case 1996). This is an individually administered oral test that takes about 10 minutes to complete, and was conducted by trained staff from a specialist data collection organisation. This test has been identified as highly predictive of achievement in primary mathematics (Gersten, Jordan and Flojo 2005). Moreover, Cowan (2011) found this test to have high reliability, with little evidence of either

floor or ceiling effects. Further information on the Number Knowledge test can be found at http://clarku.edu/numberworlds/nw_TestInfo.htm).

Children's Key Stage 2 scores are used as the baseline (pre-test) scores in the secondary school trial[4]. All children in England sit Key Stage 2 exams at the end of primary school, when they are age 10 or 11. These tests were thus completed by children three months before the Maths Mastery secondary school RCT began[5]. The GL Assessment 'Progress in Maths' (PiM) 12 test **(**http://www.gl-assessment.co.uk/products/progress-maths) was used as the secondary school trial post-test to examine children's mathematics skills during one week at the end of the academic year (Monday 30th June 2014 – Friday 4th July 2014). This test was administered using paper-and-pencil tests by class teachers and took approximately one hour to complete.

Certain features of the Progress in Mathematics test have important implications for our analysis of the secondary school trial. Around 40 percent of PiM test questions were on material ***not*** covered as part of the Year 7 Maths Mastery curriculum[6]. A clear advantage is therefore that this test is not too closely aligned to the Maths Mastery intervention, and hence there is low risk of the treatment group having been 'taught to the test'. Yet it also offers the interesting possibility of looking at potential substitution effects. Specifically, two sub-scales have been created within the PiM test. One is formed of test questions closely aligned to the Maths Mastery curriculum (60 percent of all test questions asked), with the other formed of questions that were not (the remaining 40 percent of questions). (These questions were identified by ARK blind to our analysis). It is expected that children in the treatment group will do no better (and possibly worse) on test questions covering material that is not part of the Maths Mastery curriculum. In contrast, a positive treatment effect is expected on

[4] 'Key Stage 2 scores are used to control for children's pre-treatment achievement. Given that children sat these tests only 3 months before the Maths Mastery secondary trial begun, conducting a pre-test using the Progress in Mathematics assessment would have been of little additional value. Moreover, it would have significantly increased costs, and placed additional burden on schools.'

[5] Pupils took these tests after randomisation. However, as these are high stakes tests, it is unlikely that the allocation of their future secondary school would have influenced their performance. (Indeed, the pupils would be very unlikely to know that they would have been part of the Maths Mastery trial at the point they were taking the Key Stage 2 tests).

[6] Moreover, despite calculator work ***not*** being covered within the Year 7 MM curriculum, the PiM test included both a calculator and a non-calculator section.

questions where the Maths Mastery programme places more time, effort and emphasis[7]. We shall examine this possibility within our analysis.

Balance at baseline

Regarding the primary school trial, Figure 3 compares the distribution of Number Knowledge test scores at baseline for the treatment and control groups. To facilitate interpretation, we have standardised this measure to have a mean of zero and a standard deviation of one. Mean test scores are slightly higher in the treatment group (0.06 standard deviations) compared to the control group (-0.07 standard deviations), though this difference is not statistically significant at conventional thresholds (t=1.41; p=0.16). Further consideration of balance in the primary school trial is provided in Appendix C. This focuses upon school-level characteristics only. Treatment schools had slightly fewer children with Special Educational Needs, with English as an Additional Language, and who were eligible for Free School Meals. However, treatment schools also had slightly lower OFSTED (school inspection) ratings. However, none of these differences were statistically significant at conventional thresholds. Moreover, controlling for these school-level variables within our analysis does not appreciably change the results presented below. Consequently, we conclude that balance in the primary school trial is reasonable.

Table 3 provides the baseline comparison for the secondary school RCT. Figures before accounting for attrition can be found in the panel on the left; those after accounting for attrition can be found in the panel on the right. (We focus upon the 'including attrition' figures in our discussion below). Standardised Key Stage 2 mathematics test scores (the pre-test conducted approximately three months before the intervention began) have a mean of 0.022 for the control group and 0.036 for the treatment group; a small and insignificant difference of just 0.014 standard deviations (t = 0.21; p = 0.83). Similarly, there is a difference of just 0.02 standard deviations in Key Stage 2 reading test scores and Key Stage 1 Average Point Scores. There are similar proportions of children eligible for Free School Meals (26 percent versus 28 percent) allocated to treatment and control, and also of boys and girls (52 versus 49 percent). Indeed, the only statistically significant difference is the greater number of Asian children observed in treatment schools (26 percent) than in control schools

[7] Of course it is also possible that the MM programme impacts positively on fundamental mathematical understanding and hence children may do better on both parts of the test.

(13 percent). Nevertheless, there appears to be good overall balance between the treatment and control groups

<< Table 3 >>

<< Figure 3 >>

Analysis: Overall effectiveness

The impact of the Maths Mastery programme is determined by the following OLS regression model:

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

Where:

Y^{post} = Child's post-test scores

Y^{pre} = Child's baseline test scores

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

C = Baseline (pre-treatment) controls for other pupil characteristics (available within the secondary school trial only[8]).

ε = Error term

i = child i

j = school j

Note that by controlling for prior achievement, we improve statistical power and account for the modest statistically insignificant difference in prior achievement between treatment and control groups. To allow for the fact that the intervention is a school level intervention and there is clustering of pupils within schools, all reported standard errors are clustered at the school level. The coefficient of interest from equation (1) is β – which will show whether or not there is a positive effect of the MM treatment.

[8] Information on pupils' background from the National Pupil Database has been linked to the data from the secondary school trial. Unfortunately, consent could not be gained for a similar linkage to be made in the primary school trial.

Analysis: Heterogeneous effects

The model presented in Equation 1 has specified a common programme effect; that the impact of Maths Mastery will be the same across different groups of children and across different types of school. Yet, in reality, the impact of the programme may vary with factors such as school and teacher quality, and differences in implementation. Of course, our ability to detect such 'heterogeneous effects' is limited, due to the relatively small number of schools who took part. Moreover, our trial was not powered to detect an effect in any specific sub group. Nevertheless, we present indicative evidence on possible heterogeneity in two ways.

First, an interaction term is added to model 1, examining how the effectiveness of Maths Mastery varies by a well-known school inspection rating in England (OFSTED grade). Specifically, each school in England is periodically inspected by external assessors, and given a score of between one ('outstanding') and four ('inadequate'). Schools with lower scores are likely to have more problems with pupil behaviour, weaker management and lower quality teaching. A pedagogical change such as Maths Mastery may be less effective in such institutions, due to other challenges and pressures being faced. We examine this possibility within our analysis.

We also examine possible heterogeneity via quantile regression; does Maths Mastery have a bigger impact on the lower tail of the achievement distribution than at the top (or vice-versa)? Recall that a key element of the programme is the *mastery* of topics, with the whole class studying the same content area in-depth before moving on to the next topic together (i.e. there is little use of within-class differentiation of content by pupil ability). The intervention may consequently have a particularly pronounced positive impact upon the lowest achieving students, ensuring they have a firm grasp of basics. At the same time, a zero or even negative effect may be observed towards the top of the achievement distribution, as the most academically able children are not allowed to proceed as rapidly on to new material. Such effects would be missed by an investigation of mean outcomes alone. Therefore, to capture potentially important and interesting effects away from the average, we re-estimate equation 1 using quantile regression.

4. Results

Implementation of the programme

The implementation and fidelity of the intervention was assessed through complementary qualitative research, consisting of telephone interviews, focus groups and questionnaires. Further details can be found in Jerrim et al (2015). From this work, three key findings emerge.

First, it seems the majority of intervention schools did genuinely attempt to implement key aspects of the Maths Mastery programme. Indeed, increased use of 'enquiry based learning' (i.e. problem solving) was found within the treatment schools. Likewise, there was greater use of manipulatives and visual learning techniques such as bar modelling (see Figure 1); key aspects of the Maths Mastery approach. Increased collaboration amongst teachers was also found. In contrast, no equivalent changes were reported for control schools, who seemed to adhere closely to the principal of 'business as usual', though of course there was heterogeneity in what that constituted.

Second, despite this broad adherence to the core principals of the programme, there was also significant heterogeneity between schools in how the intervention was implemented. For instance, teachers in some schools made significant adaptions to the materials provided, while in others comparatively few changes were made. Schools also differed with respect to the amount and type of homework they set; some developed their own worksheets following the principals of Maths Mastery, while others did not. Another example is that one school implemented the programme using young teachers, though there is little evidence that this practise was widespread. Nevertheless, this demonstrates how specifics around the implementation of the programme did vary by school.

Finally, the qualitative research team put forward the view that '*success* [of the intervention] *seemed to depend upon prior beliefs of teachers*' (see Jerrim et al 2015). Indeed, there was much enthusiasm for the programme in some schools, with quite some resistance to change in others. This, along with differences in implementation, enhances the possibility that the programme may have been more effective in certain types of schools than in others. This provides further motivation for exploring heterogeneous effects of the intervention, as we shall investigate below.

Overall impact of the Maths Mastery programme

Table 4 presents estimates of the impact of the programme within primary schools. Results are presented for (a) all schools enrolled in the trial and (b) separately for cohorts A and B. The estimated treatment effect is approximately 0.10 standard deviations, with an almost identical figure for each of the two cohorts. This reaches statistical significance at the ten percent level ($t = 1.82$; $p = 0.07$), with the 95 percent confidence interval ranging from -0.01 to +0.21. There is thus some evidence that introducing this particular teaching method into England's primary schools has had a positive effect upon children's mathematics skills. But the reasonably wide confidence interval suggests there is also a degree of uncertainty around this result.

<< Table 4 >>

Estimates for the secondary school trial can be found in Table 5. The left-hand most column refers to results when total test scores are the dependent variable. The middle and right hand columns then divides this into performance on questions that were and were not covered within the Maths Mastery curriculum.

<< Table 5 >>

Table 5 suggests that the secondary school intervention was associated with a small increase in overall mathematics test scores (effect size = 0.06) though this did not reach statistical significance at conventional thresholds. As perhaps expected, the MM intervention did not have any impact upon children's performance on questions that covered topics outside the Maths Mastery curriculum, with the estimated treatment effect essentially being zero. Thus, despite substituting away from these areas, there is no evidence that the reduction in children's learning time had any detrimental impact upon their ability in these areas. In contrast, the treatment had a more pronounced effect upon material that was focused upon within the MM curriculum (effect size = 0.10), reaching statistical significance at the five percent level ($t = 2.15$; $p = 0.04$). This effect is of a similar magnitude to that found for overall test scores in the primary school trial (0.099 standard deviations)[9]. An over-arching summary of the two trials is presented in Table 6 below.

[9] In additional analysis, we examined whether there was an interaction between the MM treatment in secondary schools and (i) gender, (ii) Key Stage 2 (baseline) test scores and (iii) eligibility for Free School Meals. All interactions were small and did not approach statistical significance at either the five or ten percent level. This held true for both overall test scores and sub-components of the PiM test.

<< Table 6 >>

Heterogeneity by school and teaching quality

Table 7 presents results for the treatment-by-school inspection (OFSTED) interaction[10]. Panel A refers to primary schools and panel B to secondary schools. The top half of Table 7 provides the key model parameter estimates. Our discussion will focus however upon the grey shaded rows in the bottom half of the table. These provide the estimated treatment effect within each OFSTED-rating group. Note that these ratings were not available for four primary schools and three secondary schools, which have therefore been excluded from this part of the analysis.

<< Table 7 >>

Within both the primary and secondary RCTs, the point estimates suggest the intervention may have been more effective within higher quality schools. The treatment effect upon overall test scores in 'outstanding' schools was very similar in the primary (+0.116) and secondary (+0.127) trials. A similar finding holds for schools rated as 'good'; the Maths Mastery intervention led to a +0.076 standard deviation improvement in primary schools and +0.071 standard deviations in secondary schools. In additional analysis, we found similar results when the interaction with OFSTED 'teaching quality' scores were used rather than schools overall OFSTED grade. Together, this provides further evidence that the programme led to a small improvement in children's mathematics skills – at least when implemented within well-functioning schools.

However, the effectiveness of the programme in lower quality schools (i.e. those rated as 'requires improvement' or 'inadequate' by OFSTED) is in more doubt. The estimated treatment effect in 'requires improvement' / 'inadequate' primary schools remained positive, though was rather small (+0.059 standard deviations). This is around half the impact observed in outstanding schools (+0.116 standard deviations), though our limited sample size means this difference did not reach significance at conventional levels.

A similar finding holds for the secondary school trial, though the difference is much more pronounced. Specifically, the Maths Mastery treatment actually had a large, *negative* and statistically significant effect within 'requires improvement'/'inadequate' schools (-0.27

[10] Note we have combined the 'requires improvement' and 'inadequate' categories into one group, due to their limited sample sizes independently.

standard deviations). In other words, the intervention group actually made less progress than children in the control group. This compares to the gains made by the treatment group amongst children attending good (+0.071) and outstanding (+0.127) schools. As demonstrated by the top half of Table 7 Panel B, this difference between school-types is statistically significant at conventional levels. This, in turn, helps to explain why the previous sub-section found Maths Mastery to have a lower overall impact in secondary schools than in primary schools. Specifically, this seems to be driven by the pronounced negative impact of the programme in a small number of 'low quality' schools.

Quantile regression results

Figure 4 presents the quantile regression results, illustrating whether Maths Mastery has different effects upon different parts of the post-test score distribution. Figures running along the x-axis refer to percentiles of the post-test, with the y-axis capturing the estimated effect size. The grey (black) line with circular (square) markers refers to the primary (secondary) school RCT.

<< Figure 4 >>

There is relatively little evidence of the programme having different effects on children with different levels of mathematics ability. The primary school effect is slightly smaller at P20 and P30 (0.07) than at P60 to P80 (≈0.10), though the magnitude of this difference is small, and insignificant at conventional thresholds. In comparison, the secondary school trial effect size is around 0.05 at every decile expect P20 (where it increases to 0.10). Similar results hold when we focus only upon those test questions covered within the Maths Mastery curriculum, and if schools with the lowest two OFSTED ratings ('requires improvement' and 'inadequate') are removed. Moreover, we have also tested for interactions between treatment status and prior achievement, and found only small and statistically insignificant results. Consequently, there is little evidence that Maths Mastery is any more or less effective for children of higher or lower mathematics ability.

Programme costs

Information on the monetary costs of schools implementing Maths Mastery have been provided by the charity responsible for delivering the programme (ARK). 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). Seven days of staff time

are required for training; one day for the headmaster, two days for the head of mathematics, and two days for two mathematics teachers. To calculate the cost of headmasters' time, we take the median point on the headmaster 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 headmaster day rate of £327[12]. We then inflate this figure by a fifth to allow for other costs not directly incorporated into headmasters' salaries (e.g. employer contributions to pensions) giving a total cost of £392[13]. Analogous calculations have been made for the head of mathematics (two days training at a final day rate of £251)[14] and the class teachers (a total of four days training at a final day rate of £141)[15]. Total training costs therefore amount to £1,460. The total annual cost to the primary school is £7,460. We then estimate the average number of pupils per primary school as 57; the number of pupils initially enrolled into the primary trial (5,108) divided the number of primary schools (90). The 'per pupil' cost of delivering the primary school intervention was therefore just £131 for the year. So long as the programme does not negatively influence any other outcome, only minimal economic returns will be needed to offset this low per pupil cost.

A similar exercise has been completed for secondary schools. The upfront cost to a school of participating in the programme is £6,000 per annum. Ten days of staff time is required for training; half a day for the headmaster; 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 schools). Day rates were calculated as above. 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 was, on average, 154 pupils per secondary school (7,712 children across the 50 initially recruited schools), meaning the per pupil cost was £50 per annum.

These figures are used to calculate the Cost Effectiveness Ratio (CER); how much does it cost to raise children's mathematics test scores by 0.01 standard deviations? Table 8 presents three different estimates, using either 'optimistic', 'baseline' or 'conservative' assumptions:

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

[12] The headmaster pay scale in England and Wales (outside of London) ranges from £107,210 to £43,232. We have assumed headmasters 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 teacher's will be on the 'main' pay scale, which ranges from £22,023 to £32,187 (median 27,105).

Optimistic

- Effect of the MM programme = 0.10 standard deviations (overall effect in primary schools)
- Cost per pupil = £50 (based upon secondary school)

Baseline

- Effect of the MM programme = 0.077 standard deviations (simple average of the overall effect across the two RCTs)
- Cost per pupil = £91 (simple average of primary and secondary school)

Conservative

- Effect of the MM programme = 0.055 standard deviations (overall effect in secondary schools)
- Cost per pupil = £131 (based upon primary school)

Using the most conservative numbers (small effect of the MM programme and high costs per pupil) it costs £24 per 0.01 standard deviation increase in children's maths test scores. The analogous 'lower bound' figure using 'optimistic' assumptions (high effect of the programme and low per pupil cost) is just £5. Although this range is quite wide (reflecting the inherent uncertainty in such analyses), this should not distract from the general message that, under all scenarios, the CER is relatively low. However, it is important to note that these figures refer to the 'average' CER only; given the potential for heterogeneous impacts discussed above, the trade-off between costs and benefits is likely vary significantly from school to school.

Moreover, there are likely to be other costs of schools moving to the Maths Mastery programme; ones that are difficult to place a monetary value upon. The accompanying qualitative research of the intervention revealed that teachers had to spend additional time changing some of their material (including children's homework). Likewise, some teachers suggested the programme increased aspects of their preparation time. Another possible cost is 'disruption'; the organisational time and hassle to successfully implement such a change within schools. Such disruption costs may be relatively low in schools that were willing participants in this trial, but could be much greater in the state school population as a whole. Finally, were the programme to be rolled out at scale, one could not rule out this having an impact upon teacher attrition and an increased need for additional professional development, due to teachers' existing knowledge and skills becoming obsolete. We are particularly

mindful here of the qualitative research which did suggest some teachers did not like the programme.

5. Conclusions

East Asian economies dominate the top of important international educational achievement rankings. Two of the most frequently asked questions by education policymakers have therefore become '*what is behind these countries phenomenal educational success*' and '*what can we in the West do to catch up*'? Although there are likely to be a wide range of explanations for these countries' success (Jerrim and Choi 2013; Jerrim 2014), the impact and implementation of 'East Asian teaching methods', often loosely and ill-defined, have particularly caught Western policymakers' attention. Yet despite this interest, there is currently little evidence as to whether the introduction of any particular East Asian teaching method would represent an improvement over the current status quo in many Western countries. This study provides evidence from two RCTs to start to fill this gap in the literature. Specifically, it provides an estimate of the impact of the 'Maths Mastery' programme – a method of teaching mathematics to school children modelled broadly on the approach used in Singapore. By combining evidence from across two Randomised Controlled Trials, we find consistent evidence of small yet positive treatment effects (particularly within higher 'quality' schools).

These findings have potentially important implications for education policy and practice. On the one hand, the small effect size suggests it is unlikely that widespread introduction of this particular programme would springboard Western countries like England to the top of the PISA educational achievement rankings. In other words, it cannot be seen as a 'silver bullet' that will guarantee a country success in mathematics. Yet this does not mean that implementing this teaching method is not a worthwhile investment to make. Even small effect sizes can be economically efficient, with the combination of several such interventions potentially having a large impact overall. Thus, although we advise policymakers that further evidence is still needed, the Maths Mastery programme nevertheless shows signs of promise, and should now be tested over a longer time horizon and a greater number of schools.

This recommendation should, of course, be interpreted in light of the limitations of this study. Four particular issues stand out. First, we have estimated the effect of a small (one year) 'dose' of the Maths Mastery programme, with our evaluation conducted after the first year it has been implemented in schools. More evidence is needed on its impact after teachers have

become more familiar with its novel approach, and after children have been exposed to the programme for a prolonged period of time. Second, impact has only been measured straight after the intervention has finished. Longer-term measurement of the *lasting* impact of this teaching method is needed. Indeed, we are not able to rule out the possibility that the small positive effect we find would 'fade out' if outcomes were measured at a later date (not directly at the end of the trial). Third is the issue of external validity; schools were purposefully recruited into the two trials and were not randomly sampled from a well-defined population. Although this limitation is common to many RCT's, further work should consider the extent to which our findings generalise to the population of England's schools. Fourth, one cannot rule out the possibility that the small positive effect we observe is due to 'Hawthorne effects'; treatment and control schools knew their assignment status which could have led to differences in their behaviour throughout the academic year and when completing the post-test. Finally, we remind the reader that both the primary and secondary school trials only had modest levels of statistical power. Future work should look at ways to improve the precision of estimates at different stages of the Maths Mastery programme (e.g. primary versus secondary school), and for different types of school, including through the use of quasi-experimental methods.

References

Boaler, Jo. 2010. *The elephant in the classroom: helping children learn and love maths*. London: Souvenir Press.

Cowan, Richard. 2011. 'The development and importance of proficiency in basic calculation.' Accessed 21/ 08/ 2014 from https://www.ioe.ac.uk/Study_Departments/PHD_dev_basic_calculation.pdf

Crossley, Michael and Kieth Watson. 2009. 'Comparative and international education: policy transfer, context sensitivity and professional development.' *Oxford Review of Education* 35(5): 633-49.

Department for Education. 2012. 'What can we learn from the English, mathematics and science curricula of high- performing jurisdictions?' *Department for Education Research Report DFE-RR178*. London (England): Department for Education. Accessed 22/07/2014 from www.gov.uk/government/uploads/system/uploads/attachment_data/file/184064/DFE-RR178.pdf

Department for Education. 2014a. 'Experts to visit Shanghai to raise UK maths standards.' *Department for Education press release*. London (England): Department for Education. Accessed 02/09/2014 from https://www.gov.uk/government/news/experts-to-visit-shanghai-to-raise-standards-in-maths

Department for Education. 2014b. 'Elite teachers travel from Shanghai for pioneering maths exchange.' *Department for Education press release*. London (England): Department for Education. Accessed 02/10/2015 from https://www.gov.uk/government/news/elite-teachers-travel-from-shanghai-for-pioneering-maths-exchange

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

Gersten, Russell; Nancy Jordan and Jonathan Flojo. 2005. 'Early identification and interventions for students with mathematics difficulties.' *Journal of Learning Disabilities* 38(4): 293-304.

Guskey, Thomas. 2010. 'Lessons of mastery learning.' *Educational Leadership* 2: 52-57.

Heddens, James. 1986. 'Bridging the gap between the concrete and the abstract'. *Arithmetic Teacher* 33(6): 14-17.

Hobbs, Graham and Anna Vignoles. 2010. 'Is children's free school meal eligibility a good proxy for family income?' *British Educational Research Journal* 36(4): 673-690.

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

Jerrim, John, Helen Austerberry, Cosette Crisam, Anne Ingold, Candia Morgan, Dave Pratt, Cathy Smith and Meg Wiggins. 2015. 'Mathematics Mastery: secondary evaluation report.' *Education Endowment Foundation research report.*

Jerrim, John and Alvaro Choi. 2014. 'The mathematics skills of school children: how does the UK compare to the high performing East Asian nations?' *Journal of Education Policy* 29(3): 349-76.

Jerrim, John. 2014. 'Why do East Asian children perform so well in PISA? An investigation of Western-born children of East Asian descent.' *DoQSS working paper 14/16*. Accessed 16/10/2014 from https://ideas.repec.org/p/qss/dqsswp/1416.html

Machin, Stephen and Sandra McNally. 2008. 'The literacy hour.' *Journal of Public Economics* 92: 1441 –62.

Micklewright, John; John Jerrim, Anna Vignoles, Andrew Jenkins, Rebecca Allen, Sonia Ilie, Elodie Bellarbre, Fabian Barrera and Christopher Hein. 2014. 'Teachers in England's secondary schools: Evidence from TALIS 2013.' Department for Education Research Report DFE-RR302. Accessed 24/10/2014 from https://www.gov.uk/government/publications/teachers-in-secondary-schools-evidence-from-talis-2013

Mulford, William; Halia Silins and Kenneth Leithwood. 2004. *Leadership for organisational learning and student outcomes.* Springer.

Okamoto, Yukari and Robbie Case. 1996. 'Exploring the microstructure of children's central conceptual structures in the domain of number.' *Monographs of the Society for Research in Child Development* 61(1-2): 27–59.

Skemp, Richard. 2006. 'Relational understanding and instrumental understanding.' *Mathematics Teaching* 12(2) 88-95.

Sowell, Evelyn. 1989. 'Effects of manipulative materials in mathematics instruction'. *Journal for Research in Mathematics Education* 20(5): 498-505.

Table 1. A comparison of England and Singapore's education systems

	England	Singapore
Average PISA test scores		
Mathematics (mean)	495	573
Reading (mean)	500	542
Science (mean)	516	551
Average class size (lower secondary school)		
Mean	24	36
Pupil:Teacher ratio		
Mean	13.2	13.7
% of GDP spent on education by government		
Percent	5.8	2.9
Average weekly working hours of teachers		
Total time at work (mean)	46	48
Total time actually teaching	20	17
Hours spent on out-of-school tuition per week		
Mean (median) hours	8.5 (6)	16 (14)
Headteachers reporting inadequate school resources		
Percent	78	35
Do teachers believe profession valued by society?		
Percent	35	68
Proportion of class time spent on maintaining discipline		
Percent	11	18
Teacher report of noise / disruption in classroom		
% agree that noise / disruption is a problem	22	36
Percent teachers who give different work to children of different ability		
Percent	63	21
Children's 'work ethic' in mathematics		
Standardised scale (mean)	-0.01	0.08
Children's 'perseverance' in mathematics tasks		
Standardised scale (mean)	-0.02	0.17
Children's 'motivation' in mathematics		
Standardised scale (mean)	-0.01	0.08

Source: Figures based upon PISA 2012, TALIS 2013 (Micklewright et al 2014) and World Bank (http://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS) data.

Table 2. A comparison of demographic characteristics and prior achievement of Maths Mastery participants to the England state school population

(a) Primary School Trial

	Cohort A		Cohort B	
	Trial schools	**England population**	**Trial schools**	**England population**
FSM				
No %	69	80	74	82
Yes %	31	20	26	18
Gender				
Female %	50	49	49	49
Male %	50	51	51	51
Month of Birth				
January %	8	8	8	8
February %	8	8	8	8
March %	8	8	8	8
April %	8	8	7	8
May %	8	9	8	9
June %	8	8	9	8
July %	8	9	9	9
August %	9	9	8	9
September %	9	9	8	9
October %	8	9	9	8
November %	10	8	8	8
December %	8	8	9	8
Foundation stage profile scales				
Personal, social, emotional development	-0.11	0	0.13	0
Communication, language and literacy	-0.04	0	0.07	0
Problem solving, reasoning and numeracy	-0.13	0	0.08	0
Knowledge of world	-0.17	0	0.10	0
Physical development	-0.13	0	0.11	0
Creative development	-0.07	0	0.15	0
Pupil n	**2,162***	**616,014**	**2,880***	**641,871**

Notes: All foundation stage profile scales have been standardised to have a mean of 0 and a standard deviation of 1. * By pupil number indicates that sample size is slightly different to figures reported in the main text. This is due to these figures being based upon administrative records, while trial data was collected directly from schools, and based upon children who were present on the day of the pre-test.

(b) Secondary School Trial

	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
Mean (SD) KS1 total points score	14.6 (3.5)	15.3 (3.6)
Mean (SD) KS2 mathematics score	68.4 (20.9)	70 (21)
Mean (SD) KS2 reading score	31 (10.2)	33 (10)
Pupil n	**7,712**	**531,145**

Notes: Authors' calculations using the National Pupil Database. Trial participants refers to pupils within the 50 schools initially recruited into the Maths Mastery secondary school trial. 'England' refers to the state school population of England as a whole. All percentages refer to column percentages. Key Stage 1 and Key Stage 2 test scores kept in original metric (i.e. they have not been standardised).

Table 3. Balance between treatment and control groups: Maths Mastery secondary school trial

	As randomised			Including attrition		
	Control	**Treat**	T - C	**Control**	**Treat**	**T - C**
Eligible for FSM						
No %	72	70	-2	74	72	-2
Yes %	28	30	2	26	28	2
Gender						
Female %	46	49	3	48	51	3
Male %	54	51	-3	52	49	-3
Ethnic Group						
White %	50	47	-3	51	47	-4
Asian %	13	26	13*	13	26	13*
Black %	21	16	-5	21	17	-4
Mixed %	8	7	-1	7	6	-1
Chinese %	0	0	0	0	0	0
Other / unclassified %	7	4	-3	7	4	-3
Standardised KS1 APS	0.007	-0.006	-0.013	0.041	0.065	0.024
Standardised KS2 mathematics score	0.014	-0.013	-0.027	0.022	0.036	0.014
Standardised KS2 reading score	-0.001	0.001	0.002	0.025	0.045	0.020
School n	**25**	**25**		**21**	**23**	
Pupil n	**3,708**	**4,004**		**2,687**	**3,251**	

Notes: Authors' calculations using the National Pupil Database. KS1 APS and KS2 scores have been standardised to have a mean of 0 and standard deviation of 1 (across pupils within the 50 schools as initially randomised). Figures reported for children where data available. * and ** indicate significant difference between treatment and control groups at the 10 and 5 percent level respectively.

Table 4. The impact of the Maths Mastery primary school programme on children's 'Number Knowledge' maths test scores

	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 score (standardised to mean 0 and standard deviation 1). Treatment effect presented in the column labelled 'Beta.' SE stands for standard error. * and ** indicate statistical significance at the 10 percent and 5 percent levels. Standard errors clustered at the school level.

Table 5. The impact of the Maths Mastery secondary school programme on children's 'Progress in Maths' test scores

	Total Progress in Maths score		Not covered in MM		Covered in MM	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**
Intervention Group (Ref: Control)						
Treatment	0.055	0.046	-0.003	0.041	0.100**	0.047
Baseline (pre-treatment) controls						
Key Stage 1 maths score	Yes		Yes		Yes	
Key Stage 1 average points score	Yes		Yes		Yes	
Key Stage 2 maths score	Yes		Yes		Yes	
Key Stage 2 English score	Yes		Yes		Yes	
Free School meal eligibility	Yes		Yes		Yes	
Ethnic group	Yes		Yes		Yes	
Gender	Yes		Yes		Yes	
English as an Additional Language	Yes		Yes		Yes	
N	**5,919**		**5,888**		**5,884**	

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

Table 6. Summary of results

	Primary school	Secondary school	Meta (combined)
Number of schools	90	50	140
School response rate	92 per cent	88 per cent	-
Number of pupils	5,108	7,712	12,820
Pupil response rate	82 per cent	77 per cent	-
Effect size	0.099*	0.055	0.077**
Standard error	0.054	0.046	0.036
95% confidence interval	-0.009 to 0.207	-0.037 to 0.147	0.007 to 0.147

Notes: Authors' calculations. Meta-analysis has given the primary and secondary school trials equal weight. Overall test scores (pre-specified primary outcome) have been reported for both trials. All standard errors clustered at the school level. * and ** indicate statistical significance of effect sizes at the 10 per cent and five per cent levels respectively.

Table 7. Heterogeneity effects by school quality

(a) Primary school

	Overall OFSTED score	
	Beta	**SE**
Intervention Group (Ref: Control)		
Treatment	0.116	0.094
Ofsted rating (Ref: Outstanding)		
Good	0.019	0.080
Requires improvement / inadequate	-0.102	0.087
Intervention*Ofsted rating		
Treatment * Good	-0.040	0.122
Treatment * Requires improvement / inadequate	-0.057	0.210
Estimated treatment effect		
'Outstanding' schools	0.116	0.094
'Good' schools	0.076	0.077
'Requires improvement' / 'inadequate' schools	0.059	0.190
Number of schools (clusters)	**3,941**	
Number of pupils	**79**	

(b) Secondary school

	Total Progress in Maths score		covered in MM		Not covered in MM	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**
Intervention Group (Ref: Control)						
Treatment	0.127**	0.058	0.185**	0.062	0.069	0.067
Ofsted rating (Ref: Outstanding)						
Good	-0.006	0.050	0.021	0.053	0.003	0.045
Requires improvement / inadequate	0.228**	0.044	0.223**	0.050	0.217**	0.040
Intervention*Ofsted rating						
Treatment * Good	-0.055	0.087	-0.069	0.090	-0.077	0.086
Treatment*Requires improvement / inadequate	-0.398**	0.100	-0.422**	0.106	-0.356**	0.094
Estimated treatment effect						
'Outstanding' schools	0.127**	0.058	0.185**	0.062	0.069	0.067
'Good' schools	0.071	0.064	0.115*	0.065	-0.008	0.052
'Requires improvement/inadequate' schools	-0.271**	0.083	-0.238**	0.087	-0.287**	0.068
Number of schools (clusters)	**41**		**41**		**41**	
Number of pupils	**4,882**		**4,882**		**4,886**	

Notes: Authors' calculations. * and ** indicate statistical significance at the 10 percent and 5 percent levels. Standard errors clustered at the school level. Grey shaded scales refers to estimated treatment effect by each OFSTED rating group.

Table 8. Estimated Cost Effective Ratio (CER) of the Maths Mastery programme

Assumption	Effect of MM programme	Per pupil cost (£)	CER: Cost per 0.01 SD improvement
Optimistic	0.099	£50	£5
Baseline	0.077	£91	£12
Conservative	0.055	£131	£24

Notes: Authors' calculations. See section 4 for discussion of how the 'optimistic', 'baseline' and 'conservative' assumptions have been set. Effect of the MM programme given in terms of standard deviations (effect sizes). The final column provides the CER – the cost of increasing pupils' maths test scores by 0.01 standard deviations.

Figure 1. An example maths question and the Maths Mastery route to the solution

Question: There are three consecutive numbers that add up in total to 42. What are these numbers?

'Standard' approach (trial and improvement): Children start with what they believe a reasonable estimate of the answer to be (e.g. 7, 8 and 9). They then find these sum up to 24, and so realise the set of numbers must be higher. Three higher numbers are therefore tried (e.g. 15 + 16 + 17), which in this example sum up to 48. Children will then add together another set of numbers, higher in value than the first set, but lower in value than the second set. This iterative process continues until they reach the answer of 13, 14 and 15.

'Maths Mastery' approach: The Maths Mastery approach involves 'bar-modelling' (shown below). Children would draw out the bars shown below or make them out of play blocks. They would then recognise that the total 'without the ones' is 39 (i.e. that 42 – 3 = 13). From this, they would then deduce that the grey portion of each bar is worth 13 (i.e. 39 ÷ 3 = 13). They would then simply 'add the ones' back on to the lower two bars to reach the answer of 13, 14 and 15.

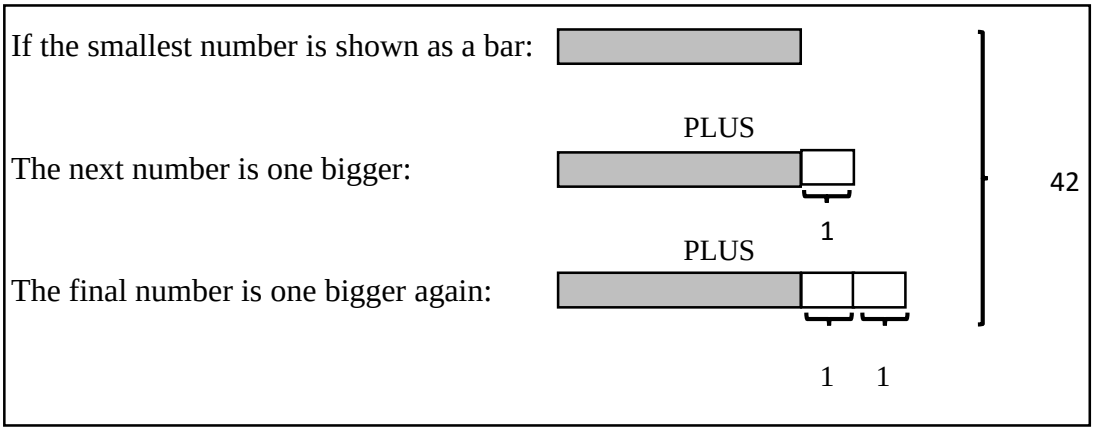


Figure 2. Support given to schools during the 'moving to mastery' year

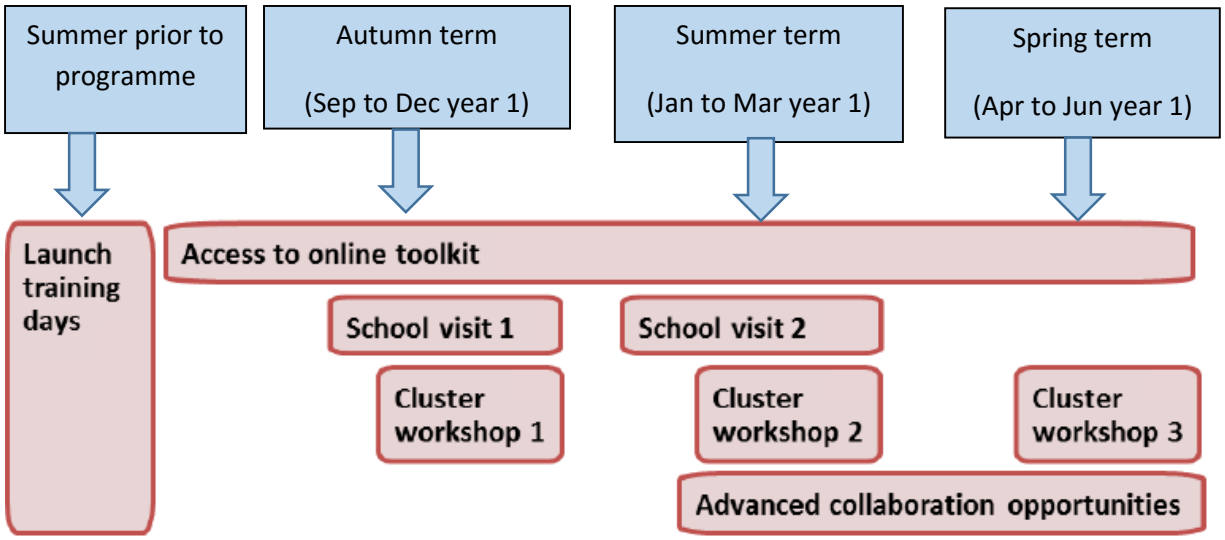
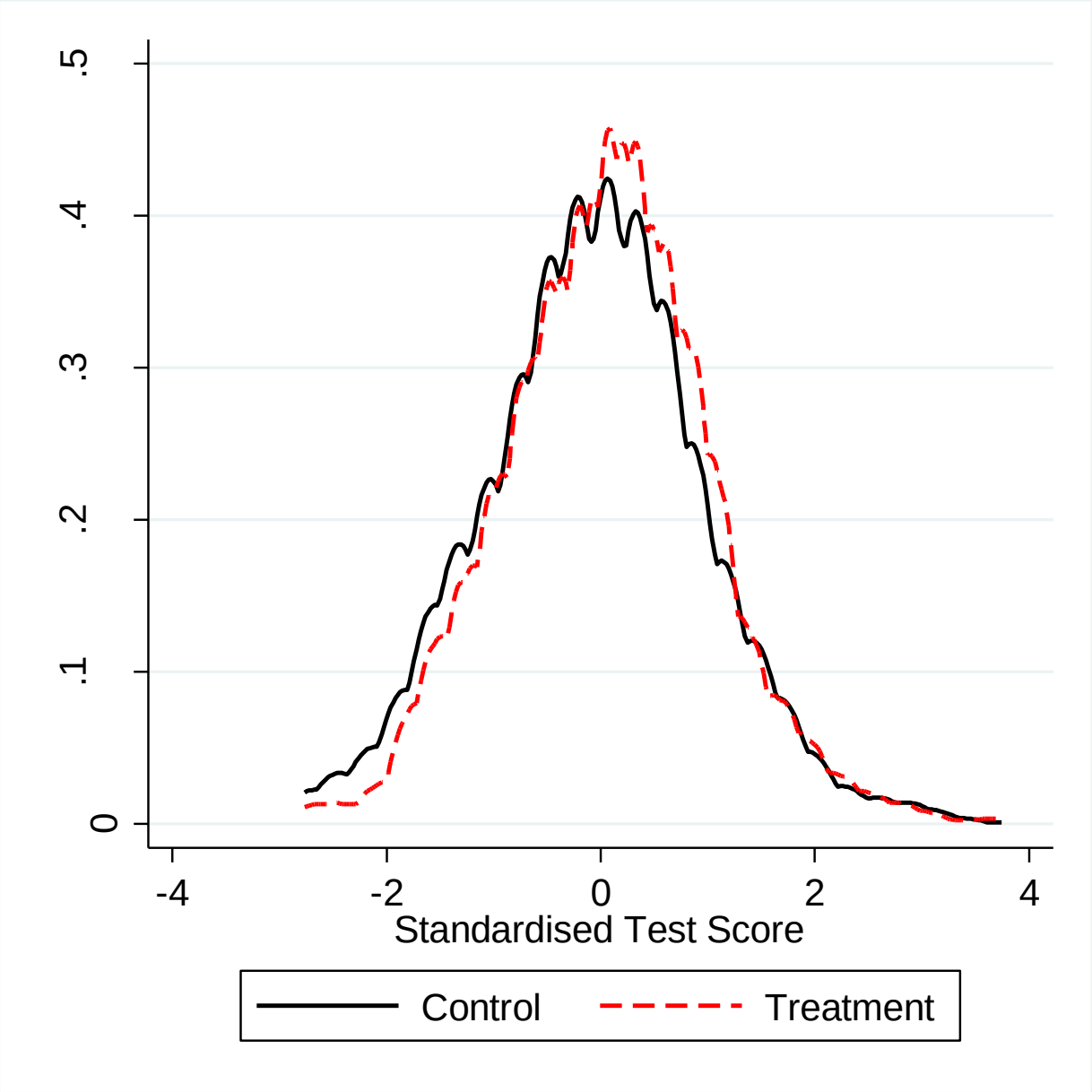
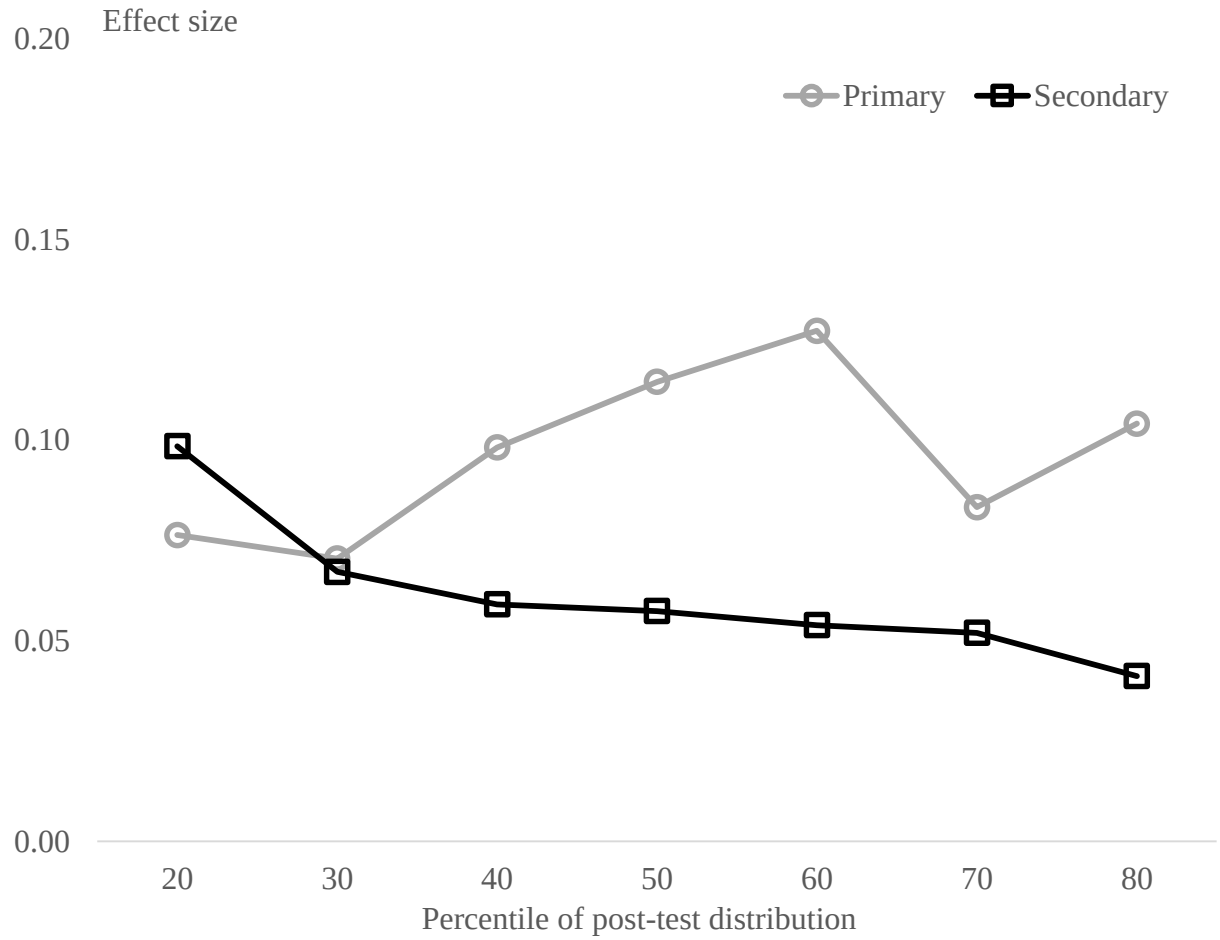


Figure 3. A comparison of (standardised) baseline test scores between the primary school treatment and control groups



Notes: Authors' calculations. Solid black line presents the baseline test score distribution for the control group. The dashed red line refers to the treatment group. Mean (median) scores equal -0.07 (0.06) for the control group and 0.06 (0.06) for the treatment group.

Figure 4. Quantile regression estimates of the effect of the Maths Mastery intervention across the post-test score distribution



Notes: Authors' calculations. Figures refer to quantile regression estimates of the impact of the Maths Mastery programme at different deciles of the post-test score distribution. Secondary school results refer to estimates based upon total scores on the Progress in Maths test.

Appendix A. Attrition from the Maths Mastery primary school trail

Appendix Table A1 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 their being slightly less evidence of selectivity in the former than the latter. Appendix Table A1 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.

Appendix Table A1. 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).

Appendix B. Attrition from the Maths Mastery secondary school trail

The National Pupil Database can be used to compare the characteristics of respondents and non-respondents across the treatment and control groups. Results are presented in Appendix Tables B1 and B2. 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 treatment group. For instance, non-respondents from the treatment 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 treatment 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, Appendix Table B2 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 percent of treatment group non-respondents were eligible for FSM, compared to just 28 percent of respondents. Likewise, 52 percent of control group respondents were male, compared to 58 percent of non-respondents.

Together, Appendix Tables B1 and B2 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. It will therefore be important to compare balance of observable characteristics between treatment and control groups both before and after attrition has been taken into account.

Appendix Table B1. A comparison of prior achievement between children who did and who did not complete the post-test

	Treatment		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.

Appendix Table B2. A comparison of demographic characteristics between children who did and who did not complete the post-test

	Treatment		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.

Appendix C. Balance of school-level characteristics in the Maths Mastery primary school RCT

	Control	Treat	T - C	Statistically significant?
% eligible for Free School Meals	30.7	26.0	-4.7	No
Key Stage average points score: Mean (SD)	29.0 (1.2)	29.1 (1.2)	0.1	No
% with English as Additional Language	60.1	54.0	-6.1	No
% with Special Educational Needs (SEN)	13.9	12.3	-1.6	No
% Outstanding OFSTED rating	33.3	25.3	-8.0	
% Good OFSTED rating	53.9	64.7	10.8	No
% requires improvement / inadequate OFSTED	12.9	10.0	-2.9	