Mathematics Mastery evaluation, GCSE attainment, 2018-19

Report from FFT Education Datalab to Mathematics Mastery

Natasha Plaister

Dave Thomson

June 2021

1 Executive summary

1.1 Methodology

- This report evaluates the effect of signing up to the Mathematics Mastery secondary programme, as measured by GCSE mathematics grade.
- We use an *intention-to-treat* analysis: we evaluate the effect on all schools that signed up to participate, regardless of the extent to which they implemented the programme in practice.
- We look at the effect on the outcomes in the 2017/18 and 2018/19 academic years.
- Our analysis used data from the National Pupil Database (NPD) to compare the performance of pupils in schools that took part in the project to the performance of pupils in a matched group of comparison schools.
- Multilevel regression models were fitted to the data, with an indicator to flag whether a school had taken part in the project. The models were adjusted to take account of pupil-level characteristics.

1.2 Main findings

- This evaluation found conclusive evidence to show that Mathematics Mastery had a positive impact on GCSE mathematics grade in 2018/19. The estimated impact on pupils in the treated group of schools was the equivalent of around 1/8 of a grade, or one month of additional progress.
- We did not find conclusive evidence to show whether or not Mathematics Mastery had a positive impact on GCSE grade in 2017/18.

1.3 Limitations

- Ideally, from an evaluation perspective, schools would have been randomly assigned to a treated group or a comparison group. As this was not the case, we constructed a comparison group of schools similar to the schools that signed up.
- Creating a comparison group in this way means that we were unable to control for factors not observed or recorded in our data.
- This evaluation was completed using *intention-to-treat* analysis. A more detailed analysis looking at how effects varied for schools depending on how closely they followed the intended programme would have been useful. However, Mathematics Mastery was only able to provide fidelity ratings, an indication of whether the school implemented the project as intended, for just over half of the participating schools. This did not give us sufficient data with which to provide an analysis of impact by level of fidelity.
- The sample size for 2017/18 outcomes was relatively small (28 schools); a larger sample may have provided more reliable and precise estimates.
- A number of Mathematics Mastery schools were newly opened at the time they joined the project, so historical data on attainment was unavailable. These schools were matched to other newly opened schools without historical data, which may mean that the comparison group is not as well-matched as it would be if historical data was available.

- Some comparison schools may have taken part in similar projects. If this improved outcomes in comparison schools, it may have led to underestimation of effects.
- We would be tentative in asserting that the results of this evaluation represent the true size of the project's impact given these limitations.

2 Introduction

Mathematics Mastery is a national programme that works with primary and secondary schools to deliver a school improvement programme. This programme is designed to empower and equip teachers to improve pupils' enjoyment, resilience, understanding and attainment in maths.

This report evaluates the impact of signing up to the Mathematics Mastery secondary programme on GCSE mathematics grade, in two outcome years: 2018 and 2019.[1] As this evaluation uses an intention-to-treat analysis, we evaluate the effect on all schools that signed up to the programme and had access to Mathematics Mastery materials, regardless of the extent to which schools actually implemented the programme in practice, if at all.

2.1 Modelling framework

This evaluation used what is known as a quasi-experimental design. This involves comparing the outcomes of pupils who went to a school that took part in Mathematics Mastery to those of pupils from a matched comparison group of statistically similar schools. This approach mimics what would be done in a formal experiment such as a randomised control trial.

We selected schools that were similar with respect to:

- Three years of historic GCSE attainment data in maths (where available)
- Three years of historic GCSE overall attainment (e.g. Attainment 8, where available)
- % pupils eligible for free school meals
- % pupils with a first language other than English
- Mean prior attainment at KS2 for each cohort on roll

A number of schools did not have three years of historic attainment data available; these were relatively new schools that did not have any pupils completing the assessments in the relevant years. These schools were matched separately, without historic attainment data.

Only mainstream state-funded schools in England were considered for the comparison group. We also excluded selective schools, as none of the treated schools were in this category. The comparison group was constructed using a technique known as covariate balancing propensity score matching (CBPS)[2]. This procedure reweights the non-participating schools such that, in aggregate, the set of non-participating schools and the set of participating schools balance (i.e. match) on the pre-treatment characteristics listed above.

Having calculated weights for each school, we estimated the effect of the programme using weighted pupil-level data. We used multilevel models, nesting pupils within schools and controlling for a number of confounding variables related to the outcome: gender, whether their first language was English, whether they were eligible for free school meals

---

[1] Throughout this report, years refer to the year in which the academic year finished – that is, 2018 refers to 2017/18, for example.

[2] See Imai and Ratkovic, 2014: https://imai.fas.harvard.edu/research/files/CBPS.pdf

and prior attainment. Confidence intervals were obtained for our estimates by using bootstrapping.

2.2 Data

Mathematics Mastery provided a dataset consisting of all schools in England that took part in Mathematics Mastery's secondary programme during the relevant time period. This included the name of the school, its URN, and the years in which it participated in the programme. In addition, a fidelity rating (low/ high) was provided where available. However, in practice the fidelity rating was only available for just over half of the schools included in this evaluation, which did not give us sufficient data with which to provide an analysis of impact by dosage.

This data was linked to corresponding records in the National Pupil Database (NPD), and to publically available school-level data. The NPD is an administrative data resource maintained by the Department for Education and provides a history of enrolments, attendance, exclusions and attainment in national tests and public examinations (e.g. GCSE and A-level) for all pupils who have been in state-funded education since 2002. For this project, we used data on attainment in GCSEs, as well as prior attainment during Key Stage 2. We also used some additional demographic variables.

All of the schools in the original dataset were successfully matched to records in the NPD. Two of the Mathematics Mastery schools merged in 2019, and were treated as one school for the purposes of this analysis. A small number of newly opened schools were excluded from the analysis because they had insufficient data available for the matching process to be carried out. Table 1 shows the total number of schools in the original and final datasets.

Table 1: Mathematics Mastery schools included in the analysis

Year	No. schools originally submitted	No. schools used in final analysis
2018	28	25
2019	36	31

This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

3 Mitigation of confounding effects

In this section, we will start with an overview of how the Mathematics Mastery schools compared to the potential comparison schools. We will then discuss how successful our chosen technique was in creating a matched control group.

From this point onwards, we will refer to schools that took part in Mathematics Mastery as *treated schools* and all other state-funded, mainstream schools as *potential comparison schools*.

3.1 Differences between treated and potential comparison schools

Overall attainment in treated schools tended to be lower than that in potential comparison schools, both overall and in maths. Schools that joined the project in 2013 had an average (capped) point score per pupil was 42.6 in the year before joining, compared to 43.8 for all other schools. For schools that joined in 2014, the average was 42.7 in the year before joining, compared to 43.6 in all other schools.[3] In 2013 joiners, just 64.6% achieved grade C or above in GCSE maths in the year before joining the project, compared to 71.6% of all other schools, and similarly for 2014 joiners, 67.1% achieved grade C or above, compared to 73.6% in all other schools.

Treated schools also tended to have lower prior attainment at Key Stage 2. For 2013 joiners, the average point score in the year they joined the project was 26.6 compared to 28.0 for all other schools, and for 2014 joiners it was 26.8 compared to 27.9 for all other schools. As for treated KS2 schools, treated KS4 schools had a high proportion of pupils who were eligible for free school meals. In 2013 joiners, 25.5% of pupils were eligible in the year they joined the project, compared to 15.3% of pupils in all other schools; in 2014 joiners, 27.4% were eligible compared to 14.7% in all other schools. They also had a high proportion of pupils with English as an additional language; 24.3% in the year of joining the project for 2013 joiners, compared to 13.5% of all other schools, and 27.4% for 2014 joiners, compared with 14.9% of all other pupils.

3.2 Extent of success in creating matched controls

The graphs in figure 1, known as love plots[4], show how similar the treated and comparison schools were to one another, before and after matching, using a measure called the standardised mean difference. The mean difference is simply the difference between the average value of the variable for the treated schools, and the average value for the comparison schools. Standardising this measure, which is done by dividing it by its standard deviation, means that we can compare balance across different variables. Generally, a standardised mean difference of 0.2 or below is considered to indicate good balance. This threshold is shown on the graphs as a dotted line.

---

[3] At this time, this measure was calculated as the average of the point score for a pupil's best eight GCSEs, with grades translating to points as follows: A* = 58, A = 52, B = 46, C = 40, D = 34, E = 28, F = 22, G = 16.

[4] Loveplots are named for Professor Thomas E. Love, who first developed them along with colleagues (https://academic.oup.com/eurheartj/article/27/12/1431/647407)

As shown in figure 1, we were able to create a very closely matched comparison group; all of the standardised mean differences are well below the 0.2 threshold after matching.

Figure 1: Loveplots showing balance before and after matching

4 Results

This sections gives the estimated impact of signing up to the Mathematics Mastery secondary programme on the two outcomes in each of the outcome years.

4.1 Formatting of results

Estimated impact is given in the same units as the outcome measure. In this report, the outcome measure is GCSE mathematics grade. An estimated impact of one on grade GCSE grade would mean that we'd expect a pupil from a Mathematics Mastery school to attain one grade higher than an equivalent pupil from a comparison school.

However, when using estimated impact it is difficult to compare across different outcome measures. An estimated impact of 0.75, for example, on GCSE grade is not the equivalent of an estimated impact of 0.75 on another outcome, such as Key Stage 2 scaled scores in maths. This means that it would be difficult to compare the effect of Mathematics Mastery's programme for secondary schools to another programme that focuses on a difference outcome measure using estimated impact.

The effect size is used to get around this problem. It is a standardised version of the estimated impact. That is, it is the estimated impact divided by the standard deviation in the outcome measure among all pupils entered for a particular subject. Because it is a standardised measure, it can be compared across different outcomes.

However, effect sizes can be difficult to interpret; it is not immediately obvious whether an effect size of, for example, 0.5 is large or small. Months of progress are a measure used in education research to try and help with this. In this report, effect sizes were translated into equivalent months of progress using guidance developed by the Education Endowment Foundation[5], as shown in table 2. In our example, an effect size of 0.5 would be the equivalent of six months of additional progress; expressed using the months of progress measure, it is clear that this is a large effect.

The use of effect sizes and months of progress in education research is not without limitations. Most importantly for this evaluation, effect sizes for interventions among primary-age children tend to be larger. This is because there is a greater spread of attainment among older children.[6] This may be particularly relevant for a project such as Mathematics Mastery that works with pupils from primary to secondary age when attempting to compare the effect of the project on different age groups.

Table 2: Effect sizes and equivalent months of progress

Effect size from	To	Months of progress

[5] As described at https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/evaluator-resources/writing-a-research-report, accessed January 2020

[6] See Higgins and Katsipataki (2016), pp. 16-19 for an overview of the criticism of effect sizes and months of progress, as well as a summary of the benefits

-0.04	0.04	0
0.05	0.09	1
0.10	0.18	2
0.19	0.26	3
0.27	0.35	4
0.36	0.44	5
0.45	0.52	6
0.53	0.61	7
0.62	0.69	8
0.70	0.78	9
0.79	0.87	10
0.88	0.95	11

4.2 GCSE mathematics grade

GCSE grades are shown in this section as point scores ranging from 9-1, with a difference of one point being the equivalent of one grade. An estimated effect of 0.5, for example, would suggest that pupils in Mathematics Mastery schools achieved the equivalent of half a grade more than pupils in comparison schools, after controlling for pupil demographics.

Estimates of the impact of Mathematics Mastery on attainment in GCSE maths are shown in table 4, with 95% confidence intervals (all to two decimal places). Results are also summarised in figure 3.

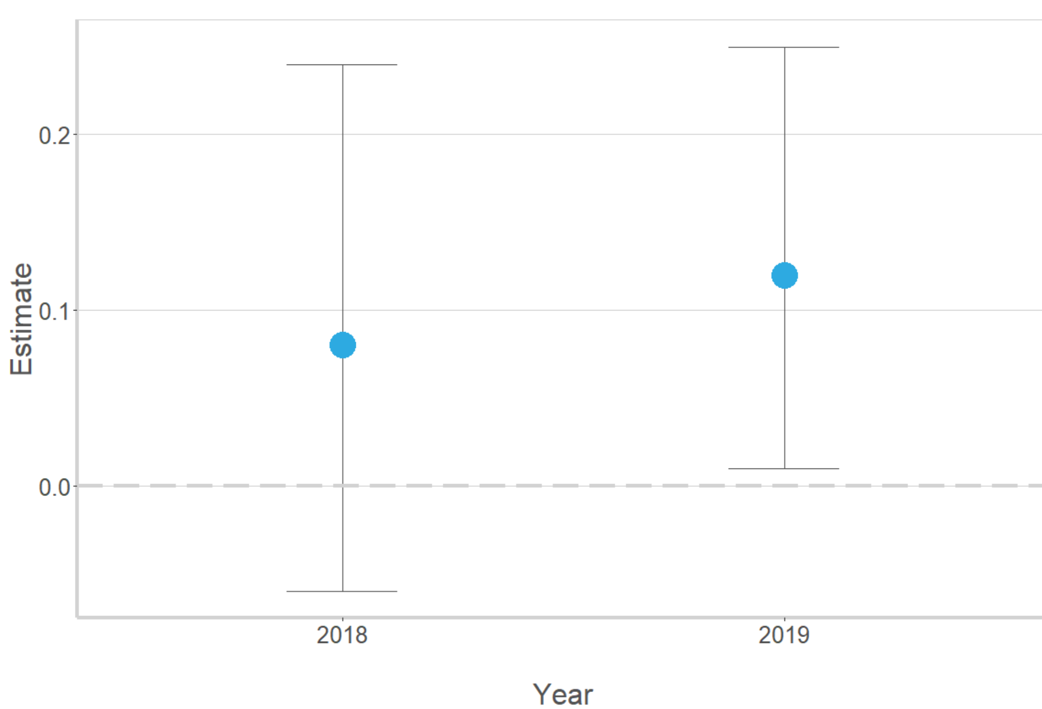
Table 4: Estimated effect of Mathematics Mastery on GCSE mathematics grade

Year	Lower CI	Estimate	Upper CI	Effect size	Months of progress	No. schools	No. pupils
2018	-0.06	0.08	0.24	0.04	0	25	7,696
2019	0.01	0.12	0.25	0.06	1	31	9,562

These results provide evidence that Mathematics Mastery had a positive impact on attainment in GCSE mathematics in 2019. We would estimate that a pupil in a Mathematics Mastery school would achieve a grade 0.12 higher than a pupil in a matched comparison school, after controlling for pupil characteristics. This is the equivalent of around 1/8 of a grade, or one month of additional progress.

However, the results do not provide conclusive evidence of an impact on attainment in GCSE mathematics in 2018. The estimate for this outcome year, while positive, is small; it is not large enough to be the equivalent of any additional months of progress, and it is not significant; that is, the confidence interval contains zero. This means that we cannot be confident that there is any impact on this outcome in 2018.

Figure 3: Estimated effect of Mathematics Mastery on GCSE mathematics grade



5 Conclusions

5.1 Overview

This evaluation found evidence to show that Mathematics Mastery had an impact on GCSE mathematics grade in 2019. We would estimate that a pupil in a Mathematics Mastery school would achieve a grade 0.12 higher than a pupil in a matched comparison school, after controlling for pupil characteristics. This is the equivalent of around 1/8 of a grade, or one month of additional progress.

However, we did not find conclusive evidence to show that Mathematics Mastery had an impact on GCSE grade in 2018. The estimated impact, while positive, was not large enough to be the equivalent of any additional months of progress, and was not statistically significant.

5.2 Limitations

This evaluation matched treated schools to comparison schools using observational data from the National Pupil Database (NPD). This type of evaluation is known as a quasi-experimental design. However, ideally, from an evaluation perspective, the project would have been provided to schools as part of a randomised control trial (RCT).

With a quasi-experimental design, there are a number of possible problems. In our analysis, we had to rely on the data in the NPD, but the NPD data is limited. For example, it does not include information about social class, parental occupations or school funding levels. Not accounting for these unobserved variables may introduce bias into our estimates. Using a quasi-experimental design also leaves open the question of how schools were selected to join the project. If there were systemic differences between the Mathematics Mastery schools and control schools - for example, if the project targeted schools in which teachers had low confidence in teaching maths - then these selection effects would pose difficulties to the evaluation.

This evaluation was completed using intention-to-treat analysis; it evaluates the impact on all schools that took part in the project, regardless of how closely schools followed the intended programme. A more detailed analysis looking at how effects varied for schools depending on how closely they followed the intended programme would have been useful. However, Mathematics Mastery was only able to provide fidelity ratings, an indication of whether the school implemented the project as intended, for just over half of the participating schools. This did not give us sufficient data with which to provide an analysis of impact by level of fidelity.

The sample size for 2018 outcomes was relatively small (28 schools). Although this still gave a reasonable number of pupils on which to fit regression models, the smaller sample may mean that estimates are less reliable. A larger sample may have provided more reliable estimates and / or smaller confidence intervals.

Mathematics Mastery worked with a number of newly opened schools. This meant that historical data on attainment was not available for matching. These schools were matched to other newly opened schools using demographic data and data on prior attainment, where possible. The omission of historical attainment data from the matching process means that we are less confident that the comparison group for these schools was well-matched, and may have led to over or underestimation of impact. However, omitting newly opened schools from the analysis altogether would have further reduced the sample size.

Some comparison schools may have taken part in similar projects, or teachers from those schools may have attended training similar to that offered by the projects. If this was the case, our analysis would not be an evaluation of the project against no equivalent support, but instead against no support in some cases and other, similar support in the rest. This could lead us to underestimate the effect of the projects, assuming that the equivalent support had a positive effect on some comparison schools' outcomes. We would note, however, that not controlling for this effect may be the relevant analysis as it represents an evaluation of the project against current conditions, with schools' choices to engage with other projects or training being included in the makeup of controls.

We would be tentative in asserting that the results of this evaluation represent the true size of the projects' impact for the reasons outlined above. The ideal evaluation of the project would have come from a fully randomised control trial which would allow for isolation of project participation as a lone variable of interest. As this was not the case, the above results represent the best estimate of the effectiveness of participation in the project that we were able to provide.