The association between within-class grouping and children's achievement in mathematics during Year 2, Year 5, and Year 9.

School choices report

March 2021

Author:

John Jerrim

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

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

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

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education. Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

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

Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The study was conducted by John Jerrim, UCL Institute of Education. Email: J.Jerrim@ucl.ac.uk

Acknowledgements

The report author would like to acknowledge the useful comments received from the EEF project steering group (particularly the external members: Christian Bokhove and Richard Dorset), and anonymous peer-reviewers.

Executive summary

The project

Evidence of an association between within-class attainment grouping and pupils' achievement is relatively weak, particularly for secondary school pupils. Little research is available comparing within-class attainment grouping to alternative approaches (such as mixed-attainment grouping) other than whole-class teaching. Moreover, little is currently known about the practises and approaches actually in use in England, such as how often pupils are exposed to whole-class teaching as opposed to mixed or same-attainment grouping, and whether teachers use one single approach or a mix of these.

This study presents new evidence on these questions within England. The aim of this study was to investigate whether there is any association between within-class attainment grouping and children's academic achievement in mathematics. The study also looked at pupils' enjoyment of number work (Year 2), self-confidence, and achievement in reading as secondary outcomes. A key goal of this study was to compare two alternative teaching approaches: (a) whole-class teaching and (b) teaching in (mixed or same) attainment groups, also providing descriptive information on the extent that these different approaches are currently used by teachers in England.

The study focused specifically on Year 2 pupils across the U.K. and Year 5 and Year 9 children in England. The final analysis is based upon 7,899 Year 2 pupils, 2,856 Year 5 pupils, and 3,288 Year 9 pupils.

This is a secondary data analysis project that used available data collected as part of the Millennium Cohort Study for Year 2 pupils and the 2015 round of the Trends in Mathematics and Science Study (TIMSS) for Year 5 and Year 9 pupils. Regression analysis, controlling for children's prior achievement and a range of other characteristics, was used to explore the association between within-class attainment grouping and children's achievement.

Due to the observational nature of this study, the results presented in this report refer to conditional associations and are not intended to capture any possible causal links.

The study was conducted by John Jerrim, UCL Institute of Education between October 2019 and November 2020.

Key conclusions	
1.	There is no evidence of an association between within-class attainment grouping and Year 2 children's mathematics test scores.
2.	There is no evidence that the frequency of using any of the teaching approaches (same-attainment groups, mixed-attainment groups, or using a whole-class approach) is associated with higher mathematics achievement amongst Year 5 or Year 9 pupils (that is, none of the approaches were found to be superior to the others).
3.	There is little evidence that the frequency of using different teaching approaches (whole-class teaching, within-class same-attainment grouping, or within-class mixed-attainment grouping) is associated with Year 5 and Year 9 pupils' self-confidence in their mathematics abilities.

Additional findings

Overall, there is no clear evidence that any of the three teaching approaches (whole-class teaching, within-class mixed-attainment grouping, and within-class same-attainment grouping) were associated with higher (or lower) mathematics achievement, nor that any of the these approaches were particularly beneficial for disadvantaged or low-achieving pupils' progress in mathematics in Year 5 or Year 9.

The only exceptions, for which evidence continues to remain weak, include the use of within-class attainment grouping in Year 2 and whole-class teaching in Year 9.

For Year 2 pupils from low-income backgrounds, the use of within-class attainment grouping was associated with a 0.14 standard deviation increase in their mathematics achievement. However, the same finding does not emerge for the secondary outcomes (enjoyment of number work and reading test scores). Similarly, children with high levels of prior achievement seem to benefit, on average, from the practise of attainment-grouping in mathematics (effect size 0.10), though again the same pattern does not emerge for either of the secondary outcomes.

Disadvantaged and low-achieving Year 9 pupils who received frequent whole-class teaching did experience slightly faster growth in their mathematics achievement although there is quite a large degree of uncertainty surrounding this result and the effects appear to be very small.

There is little evidence that the use of within-class attainment grouping of Year 2 pupils is strongly associated with common observable teacher and pupil characteristics. However, there seem to be varying levels of correlation (for Year 5 and Year 9) between teachers' years of experience, the type(s) of teaching approach(es) chosen, and the frequency of their use. For instance, less experienced teachers in Year 5 seemed less likely to teach in mixed-attainment groups whilst less experienced teachers in Year 9 were slightly less likely to use whole-class approaches.

Within-class attainment grouping was found to be more common for Year 2 pupils in schools where setting or streaming of pupils was not used. For Year 5 pupils, within-class mixed-attainment grouping occurred less frequently than either same-attainment grouping or whole-class teaching. For Year 9 pupils, whole-class teaching was the most common approach used to teach mathematics while within-class mixed-attainment grouping of pupils was comparatively rare.

The null results of this study show that evidence is not strong enough to suggest that any single approach is better than any of the others, and might also indicate that teachers use the approach (or a combination of approaches) that is most comfortable (and effective) for them and their pupils. However, the results suggest that further research investigating the link between whole-class teaching and pupils' mathematics outcomes may be worth pursuing.

Introduction

Background

In almost every education system across the world, children are separated into different 'attainment' (or achievement) groups at some stage. Yet this is done at different ages and in different ways. In some countries, children are divided into different schools based upon their achievement (between-school tracking), with a domestic example being the continuing prevalence of grammar schools within certain parts of the U.K. Alternatively, children may be divided into different attainment groups within the school that they attend. (It should be noted that the term 'attainment grouping' is used throughout this report in order to be consistent with much of the existing literature on this topic. However, 'attainment grouping' may actually be a more appropriate term given that pupils are usually separated into different groups based upon measures of their achievement in key subjects.) This can take the form of streaming (dividing children into the same-attainment group for all subjects), setting (dividing children into attainment groups on a subject-by-subject basis), or within-class attainment grouping (for example, children of similar attainment being placed on the same class table together). Of course, teachers and schools can (and do) use a combination of these approaches so that children are first 'set' for a given subject and then further divided by attainment within their class. Yet, despite the fundamental importance of this issue, little is currently known about why some teachers decide to group children by attainment within their classes and the association that this has with their subsequent academic achievement. This is particularly true within the context of the English education system, a nation where within-school attainment grouping is extensive (Education Datalab, 2019). The aim of this report is to start to address this gap in the literature by providing new evidence on the use of within-class attainment grouping within Year 2, Year 5, and Year 9 classes in England and the association that this has with children's skills (particularly in mathematics).

Why might teachers and schools choose to group pupils by attainment? Proponents of attainment grouping suggest that it allows teachers to set different tasks for different groups of pupils that most appropriately matches their existing knowledge and skills (Coe et al., 2008). This then enables teachers to stretch the highest-achieving pupils with more challenging material while lower-achieving children develop firmer foundations in the basics to ensure that they are not left behind (Francis et al., 2017). It may also enable teachers to deploy their resources in what they believe to be an effective manner (for example, to allocate class teaching assistants to provide extra support to the groups in most need). Within-class attainment grouping may also make the workload of teachers more manageable—including the management of disruptive pupils (Smith and Sutherland, 2003).

In contrast, those who argue against the use of attainment grouping suggest that there is a lack of empirical evidence to support the claims made above (Taylor et al., 2017). Rather, it unhelpfully means some children get labelled as low-achievers, leading them to develop negative attitudes towards school or a particular subject (Archer et al., 2018), with their academic achievement suffering as a result. This may be supported by evidence from the peer-effects literature, with some studies suggesting that low-achieving pupils tend to make more academic progress when they are placed in classes with higher-achieving students (Burke and Sass, 2013).

Despite the above, existing evidence on the impact of attainment grouping upon pupils' outcomes remains limited. Indeed, as the Education Endowment Foundation (EEF) toolkit notes, the existing evidence on the impact of within-class attainment grouping is 'limited'. There is a particular lack of studies comparing the impact of within-class attainment grouping to within-class grouping of pupils not based upon attainment.

This project uses data from the Millennium Cohort Study (Year 2 in 2008/2009; n = 7,899) and the Trends in Mathematics and Science Study (Year 5, n = 2,856 and Year 9, n = 3,288 in 2015) to investigate the association between within-class attainment grouping, children's achievement (primarily in mathematics), and their academic self-confidence. It uses regression analysis including a rich set of controls for children's demographic backgrounds, prior achievement, teacher, class, and school characteristics. This is deemed the most appropriate analytic method to address this question given the likely difficulty in conducting an RCT that would require teachers to artificially change their usual teaching approach. To interpret estimates from this method as causal, a 'selection upon observables' assumption must hold (that is, that the statistical models estimated include all factors that influence both the decision of teachers to use within-class attainment grouping and children's outcomes). As this is unlikely to be the case, it is more appropriate to treat the estimates presented in this report as conditional associations, rather than capturing cause and effect.

Intervention

The ‘intervention’ in this context refers to the approach used by teachers in the instruction of mathematics. A particular focus is how ‘within-class attainment grouping’ (such as pupils with similar levels of mathematics achievement being sat together on the same class table) is associated with achievement in mathematics. In the analysis of the Millennium Cohort Study (MCS), this is based upon a binary measure of whether within-class attainment grouping is used (yes/no). In the analysis of the TIMSS dataset for Year 5 and Year 9 pupils, this is compared to two alternative teaching approaches: (a) whole-class instruction and (b) teaching pupils in (within-class) mixed-attainment groups. The subsections below describe how this is operationalised within the datasets.

Millennium Cohort Study

Within the MCS, teachers were provided with definitions of within-*class* attainment grouping and within-*subject* attainment grouping (the precise wording is provided in Appendix F). They were then asked:

- In this child’s class, is there within-class attainment grouping (yes/no)?

To those teachers who selected ‘yes’, they were then asked:

- How many within-class attainment groups are there?
- Which group is this child in (highest, middle, or lowest)?

All teachers were then asked analogous questions specifically about within-subject attainment groupings in literacy and mathematics. For example:

- In this child’s class, are there within-class subject groups for literacy (yes/no)?
- How many within-class subject groups are there for literacy?
- Which group is this child in for literacy (highest, middle, or lowest)?

The MCS survey organisers have used teachers’ responses to these questions to derive a set of variables that have been deposited as part of the MCS dataset. These are listed below and are the key covariates of interest within the analysis:

- wicagn: the within-class attainment group of the child in mathematics (not grouped, top group, middle group, or bottom group); and
- wicagl: the within-class attainment group of the child in literacy (not grouped, top group, middle group, or bottom group).

Specifically, based upon these variables, the analysis investigates the ‘impact’ of being (a) grouped within class versus not being grouped within class and (b) for those who are grouped within class, the ‘impact’ of being allocated to the top, middle, or bottom group. The primary outcome of the MCS analysis is children’s mathematics achievement as measured by scores on a short-form Progress in Maths test. The secondary outcome is children’s enjoyment of number work measured using a single question included within the MCS questionnaire.

Trends in Mathematics and Science Study

As part of the TIMSS study, teachers within each school were asked to complete a questionnaire. This included the question:

In teaching mathematics to the students in this class, how often do you ask pupils to—

- work problems together in the whole-class with direct guidance from yourself?
- work in mixed-attainment groups?
- work in same-attainment groups?

There were four possible response options: 1, ‘every or almost every lesson’; 2, ‘about half the lessons’; 3, ‘some lessons’; and 4, ‘never’. Note that, to avoid small sample sizes, the bottom two categories (‘never’ and ‘some lessons’) have been combined into a single group. The primary exposure of interest in the analysis is the frequency with which mixed-attainment grouping is used. The ‘effect’ of same-attainment grouping upon children’s achievement can also be compared to the ‘effect’ of (a) mixed-attainment grouping and (b) whole-class teaching. The primary outcome of the TIMSS analysis is pupil’s scores on the TIMSS maths test. The secondary outcome is children’s self-confidence in mathematics captured by a set of questions included in the TIMSS background questionnaire. Note that this data is linked to the National Pupil Database (NPD), which provides measures of socio-economic status—free school meal (FSM) eligibility—and prior achievement (Key Stage test scores).

Evaluation objectives

The project attempts to answer the following research questions:

1. To what extent do Year 2, Year 5, and Year 9 teachers in England use within-class attainment grouping? How does this vary by school, teacher, classroom, and pupil characteristics?
2. Is within-class attainment grouping associated with higher levels of achievement?
3. Is being placed in a higher within-class attainment group associated with greater academic progress? (Year 2 pupils only.)
4. Is there an association between within-class attainment grouping and children's academic self-confidence (Year 5 and Year 9) and enjoyment of mathematics (Year 2)?
5. Does the association between within-class attainment grouping and children's achievement differ depending upon children's family background, prior academic achievement, and whether the school sets/streams pupils?

Ethics and trial registration

The project was subject to ethical review by the UCL Institute of Education's Research Ethics committee. The study received final ethical approval on 17 December 2019 (reference: REC1279 – within-class attainment grouping).

A study plan was developed, including input from an external advisory board, before data analysis was conducted. This pre-specified study plan is available from the EEF website.[1]

Data protection

Legal basis for processing personal data

UCL relies upon Article 6(1)e of the GDPR: processing is necessary for the performance of a task carried out in the public interest as the lawful basis for processing data in this project. The research project falls within UCL's research core purpose performed in the public interest, as set out in UCL's Statement of Tasks in the Public Interest.[2]

Legal process for processing special category data

Access to special category data has been requested under Article 9(2)(j) of the GDPR: 'Necessary for archiving, scientific, historical research or statistical purposes.' Within Schedule 1, Part 1 of the DPA 2018, this project meets condition 4: it is necessary for scientific research purposes.

The research question this project answers represents a pressing social need. It is important that education policy and practise are evidence led. Yet we currently know little about the association between how teachers choose to structure their classrooms and children's academic outcomes. This includes within-class attainment grouping, where children of different perceived abilities are separated into different groups (for example, sit at different tables) depending upon the choice made by their teachers. Such practises could potentially be associated with both children's achievement and levels of educational inequality. Yet we currently know very little about this issue. This project, including having access to special category data, provides new evidence on this socially important matter.

Special category data is used in this project as this information needs to be controlled for within the statistical analysis. For instance, it is plausible that children with certain characteristics are more likely to be placed into one mathematics group rather than another by teachers. It is hence necessary to control for such characteristics within the analysis

[1] https://educationendowmentfoundation.org.uk/public/files/Projects/Within_Class_Grouping_in_Maths_Study_Plan.pdf
[2] https://www.ucl.ac.uk/legal-services/sites/legal-services/files/ucl_statement_of_tasks_in_the_public_interest_-_august_2018.pdf

otherwise one is not able to rule out such factors driving the results. This made access to such special category data necessary for scientific research purposes.

GDPR compliance

To ensure the appropriate safeguards are in place, this project has closely followed the 5 SAFEs Framework.

- Safe people: the report author was the sole person who had access to the relevant data; he has completed the necessary ONS training on being a safe researcher.
- Safe setting: the TIMSS-NPD analysis was conducted via the ONS Secure Remote Service; no data was therefore physically handled or transferred.
- Safe projects: this project report is made available online free of charge; the evidence from this project has the potential to help schools and teachers decide the most appropriate methods to use to teach mathematics to their pupils; it extends our knowledge of the association between within-class attainment grouping and pupil achievement, which may aid education policy decision-making.
- Safe data: to adhere to the principal of data minimisation, only a subset of available fields from the TIMSS-NPD data is used (with all fields having low levels of sensitivity and risk of identification).
- Safe outputs: this report has been subject to statistical disclosure control, with small sample sizes supressed. The TIMSS-NPD analysis has undergone checks by the ONS before it was released from the SRS.

Project team

The project was undertaken by Professor John Jerrim from the UCL Institute of Education. He completed the data analysis and writing of the research report.

Methods

Study design

Table 1: Research design

<table>
<tr><td colspan="2">Design</td><td>Regression analysis</td></tr>
<tr><td colspan="2">Unit of analysis
(school, pupils)</td><td>Pupils</td></tr>
<tr><td colspan="2">Number of children included in analysis</td><td>Year 2: 7,899.
Year 5: 2,856.
Year 9: 3,288.</td></tr>
<tr><td rowspan="2">Primary outcome</td><td>Variable</td><td>Year 2: maths attainment
Year 5: maths attainment
Year 9: maths attainment</td></tr>
<tr><td>measure (instrument, scale, source)</td><td>Year 2: Progress in Maths test, 0–28; MCS, variable: MATHS7SC.[3]
Year 5 and Year 9: TIMSS mathematics test, 0–1,000; TIMSS variable: SMMAT01–SMMAT05.</td></tr>
<tr><td rowspan="2">Secondary outcome(s)</td><td>variable(s)</td><td>Year 2: enjoyment of mathematics.
Year 5 and Year 9: confidence in mathematics.</td></tr>
<tr><td>measure(s)
(instrument, scale, source)</td><td>Year 2: single question asked in the MCS, ‘How much do you enjoy number work?’; three-point scale (converted to binary), MCS questionnaire, variable: dcsc0023.
Scores, 10–222, MCS; variable: DCWRAB00.
Year 5 and Year 9: Students Confidence in Mathematics scale, 3–14; TIMSS questionnaire.</td></tr>
<tr><td rowspan="2">Baseline for primary outcome</td><td>variable</td><td>Year 2: the prior achievement measures included as separate covariates in the model were naming vocabulary, pattern construction, picture similarities, and Foundation Stage Profile scores (where available).
Year 5: the prior achievement measures included as separate covariates in the model were Key Stage 1 scores in mathematics and English.
Year 9: the prior achievement measures included as separate covariates in the model were Key Stage 2 scores in mathematics and English.</td></tr>
<tr><td>measure (instrument, scale, source)</td><td>Year 2
Naming vocabulary; BAS II sub-scale at age 5, 20–80; MCS, variable: cdnvtscr.</td></tr>
</table>

[3] For further details see https://cls.ucl.ac.uk/wp-content/uploads/2017/06/Data-Note-20131_MCS-Test-Scores_Roxanne-Connelly-revised.pdf

		Pattern construction scores; BAS II sub-scale at age 5, 20–80; MCS, variable: cdpctscr. Picture similarities; BAS II sub-scale at age 5, 20–80; MCS, variable: cdpstscr. Foundation Stage Profile communication, language and literacy, 0–36; MCS, variable: cll. Foundation Stage Profile mathematical development scores, 0–27; MCS, variable: md. Year 5 Key Stage 1 maths (3–39) and reading/writing (3–27) as two covariates (NPD); variables: KS1_MATPOINTS and KS1_READWRITPOINTS. Year 9 Key Stage 2 maths (0–100) and reading (0–50) scores as two covariates (NPD); variables: KS2_MATTOTMRK and KS2_READMRK.
Baseline for secondary outcome	Variable	As above for primary outcome.
	measure (instrument, scale, source)	

Participant selection

Millennium Cohort Study

The Millennium Cohort Study is a rich, nationally-representative, longitudinal study of U.K. children. A stratified, clustered survey design was used with geographic areas (electoral wards) selected as the primary sampling unit and then households with newly-born children randomly selected from within (see Plewis, 2004). Six sweeps have been conducted between 2000/2001 and 2015 when children were nine months, three, five, seven, eleven, and fourteen years old. Parents, children, and their teachers have been interviewed within the various sweeps. In total, 19,243 children participated in the first survey when children were nine months old (12,224 in England). This project draws upon data from the age five and age seven survey sweeps conducted in 2006 and 2008. By the time of the age seven survey, 13,857 children (8,896 in England) remained. The focus of this report is the prevalence of within-class attainment grouping at age seven, with this information provided by their Year 2 teachers (who as part of the MCS also completed a questionnaire), and the association this has with children’s achievement. A total of 8,498 teachers completed this questionnaire (5,567 in England). Moreover, a small number of children in England and Wales completed the MCS fourth wave survey when they were in Year 3 and so are excluded for the analysis (see Table 2 for further details). This part of the analysis draws upon data from across the whole of the U.K. Details on how the final analytic sample has been reached is presented in Table 2.

Table 2: Sample selection criteria to define the analytic MCS sample

	UK	England
Participated at wave 1	19,243	12,224
Participated at wave 4 (age 7)	13,857	8,896
Teacher survey completed	8,775	5,567
Within-class attainment grouping data available	8,498	5,394
Age 7 test score data available	8,382	5,336
Child in Year 2	7,899	5,026
Final analytic sample	**7,899**	**5,026**

Note: Separate figures shown for England to illustrate the extent that observations are driven by data from this country.

Trends in Mathematics and Science Study 2015

The Trends in Mathematics and Science Study was conducted in May/June 2015. The focus of this project is the data collected for England. A sample of pupils from two school year groups in England participated in TIMSS (Year 5 and Year 9), with this including a mathematics and science test, a background questionnaire answered by pupils, a questionnaire answered by teachers, and a questionnaire answered by headteachers.

In England, a total of 147 primary schools and 143 secondary schools were randomly selected to participate in the study. If schools refused to participate a substitute school could take their place. Within each of these schools, either one of two classes were then randomly chosen to participate. The final response rates in Year 5, after the inclusion of replacements, were 98% for schools and 98% for pupils; the Year 9 response rates were similar: 97% (schools) and 97% (pupils). This yielded a final total sample size of 8,820 pupils (4,006 in Year 5 and 4,814 in Year 9).

The following criteria is used to exclude a limited number of observations from the analysis where key pieces of information are missing:

- children are excluded if their teacher did not respond to the mixed-attainment grouping, same-attainment grouping, or whole-class teaching questions—a total of 657 Year 5 and 721 Year 9 children are dropped due to these criteria;
- children are excluded if NPD data is not available or cannot be merged into the TIMSS sample—a total of 242 Year 5 and 372 Year 9 children are dropped due to these criteria; and
- children are excluded if they are missing information on the additional control variables used in the TIMSS analysis—a total of 251 Year 5 and 433 Year 9 children are dropped due to these criteria.

The final analytic sample size in the headline analysis is therefore 2,856 Year 5 and 3,288 Year 9 pupils.

Outcome measures

Baseline measures

Throughout the analysis, multiple prior achievement measures are included within the regression models as separate covariates. This is done for two reasons. First, skills in different academic areas may be associated with skill development in mathematics. For instance, children who have stronger reading skills may improve their mathematics skills at a faster rate than those with weaker reading skills (over and above their prior achievement in mathematics). Secondly, all prior achievement measures are likely to be measured with error and hence can only partially account for potential confounding within the subject that they are meant to be measuring. The inclusion of additional controls for prior achievement helps to soak up such residual confounding resulting from baseline skill mismeasurement.

Millennium Cohort Study

The following prior achievement measures are included as separate covariates in the model. Further details about each measure is available in Connelly (2013).

- Naming vocabulary measured at age five using the BAS II sub-scale (variable name cdnvtscr). This is designed to measure children’s expressive verbal attainment. In the test, the child is shown a series of pictures of objects and asked to name them.
- Pattern construction measured at age five using the BAS II sub-scale (variable name cdpctscr). This is designed to measure children’s spatial problem-solving. In the test, the child is asked to replicate a design using patterned squares.
- Picture similarities measured at age five using the BAS II sub-scale (variable name cdpstscr). This test measures children’s non-verbal reasoning. In the test, the child is shown a row of four pictures and is asked to identify a further congruent picture.
- Foundation Stage Profile communication, language, and literacy scores. This is available for children in England only (variable name cll).

- Foundation Stage Profile mathematical development scores. This is available for children in England only (variable name md).

Trends in Mathematics and Science Study Year 5

The following prior achievement measures are included as separate covariates in the model:

- Key Stage 1 maths scores, entered as a set of dummy variables (variable name KS1_MATPOINTS); and
- the average of Key Stage 1 reading/writing scores (variable name KS1_READWRITPOINTS).

Trends in Mathematics and Science Study Year 9

The following prior achievement measures are included as separate covariates in the model.

- Key Stage 2 maths scores (variable name = KS2_MATTOTMRK).
- Key Stage 2 reading scores (variable name = KS2_READMRK).

Primary outcome

Millennium Cohort Study

There are two potential sources of information about cohort members' academic achievement around age seven.

The first is their Key Stage 1 scores in mathematics, which have been linked into the MCS from the NPD. These have the advantage of being widely understood measures of national achievement that are routinely used in secondary academic research. They are, however, only available for cohort members living in England and Wales. Moreover, within these two countries, some children are missing Key Stage 1 data (for example, due to non-consent or NPD linkage not being possible) and only broad levels (rather than fine-grained scores) are available within the MCS. They are also based upon teacher assessment rather than standardised tests. This data is also only available via the U.K. Data Service secure lab, with a six-month time-lag likely to gain access.

The second is cohort members' scores on a maths test taken as part of the MCS survey, the Progress in Mathematics test. This test is designed to measure children's mathematics skills and knowledge. It involves children completing a series of basic calculations covering numbers, shapes, measurement, and data handling. A specially designed version of the PiM was developed for the MCS that required children to complete fewer test items than the full PiM test. All children answered an initial set of seven test questions. They were then assigned to a set of four further easy, medium, or hard questions based upon their performance on the initial seven questions. A Rasch model has then been used to derive attainment scores for each cohort member. Although the test has been used in previous EEF evaluations,[4] there has been little work investigating the psychometric properties of the specific version used in the MCS.

The PiM test has the advantage of being available for a larger number of cohort members and was administered and marked by independent members of the MCS survey team. On the other hand, it is a relatively short assessment and is 'low-stakes' (pupils, teachers, and schools have nothing riding on the results). The correlation between scores on these tests and Key Stage 2 scores for MCS cohort members is around 0.6.

Based upon the above, performance in the Progress in Mathematics test is used as the primary outcome measure. Throughout the analysis, scores are standardised to a mean of zero and a standard deviation of one (using the standard deviation calculated using the full sample with data available). All estimates are therefore presented in terms of effect sizes.

[4] (https://educationendowmentfoundation.org.uk/public/files/Support/EEF_Research_Papers/Research_Paper_1_-_Properties_of_commercial_tests.pdf)

Trends in Mathematics and Science Study

Pupils in the Year 5 TIMSS sample complete a 72-minute paper-based test roughly split evenly between science and mathematics. The length of the test for Year 9 pupils was slightly longer (90-minutes), again split between science and mathematics.

TIMSS tests participants' knowledge, understanding, and application of an international curriculum, including aspects such as 'number', 'geometric shapes and measures', and 'data display' in mathematics. The fact that TIMSS attempts to measure children's knowledge and skills of an internationally determined curriculum means that not all of the test questions covered in the TIMSS test are also taught within any given country's national curriculum. However, this is less of an issue in England as around 90% of TIMSS mathematics questions are covered by the national curriculum. The test makes extensive use of multiple-choice items.5

TIMSS uses a multiple matrix-sampling approach to the test design meaning that each child is randomly assigned a subset of all test items. A complex latent regression model is then used by the survey organisers to derive estimates of pupils' achievement (known in the psychometrics literature as 'plausible values') based upon how children responded to the test question and their background characteristics. These plausible values are essentially multiple imputations of children's latent attainment in science and mathematics and capture the uncertainty in estimates of children's achievement in these two subject areas. Unless otherwise stated, recommended practise is followed in the use of these plausible values with models estimated five times (that is, five versions of the model are estimated with plausible value 1 used as the dependent variable in model 1, plausible value 2 used in model 2, and so-forth) and then combining these into the final estimate following Rubin's Rules. For further details about the plausible values and the TIMSS test design, see Jerrim et al. (2017) and Martin, Mullis and Hooper (2016).6

The primary outcome within the analysis is children's TIMSS scores in mathematics. The focus of this project is mathematics—rather than science—due to this subject being assessed for all pupils within the Key Stage 1 and Key Stage 2 tests (thus, critically, providing measures of prior achievement used within the analysis models, as outlined below).

Secondary outcomes

Millennium Cohort Study

The MCS age seven survey completed by children includes a range of questions. This study makes use of the question 'How much do you enjoy number work?' (relating to the primary outcome). This was answered using a three-point scale: 1, 'don't like it'; 2, 'like it a little'; and 3, 'like it a lot'.

Responses to this question are converted into binary format (0 = 'don't like it' and 'like it a little'; 1 = 'like it a lot'). One limitation of this as an outcome measure is that it is measuring the construct of interest (enjoyment of mathematics) using a single question.

Trends in Mathematics and Science Study

As part of the background questionnaire, Year 5 and Year 9 pupils were asked a range of questions about their attitudes towards mathematics concerning the nine areas listed below. Responses were given using a four-point scale—'agree a lot' to 'disagree a lot'—capturing their self-confidence in mathematics:

- 'I usually do well in mathematics.'
- 'Mathematics is harder for me than for many of my classmates.'
- 'I am just not good at mathematics.'
- 'I learn things quickly in mathematics.'

[5] Example questions are available at https://timssandpirls.bc.edu/timss2015/downloads/T15_FW_AppB.pdf
[6] Note that TIMSS samples pupils within classrooms and not teachers per se. This means that the population of interest in TIMSS – who inferences refer to – are pupils rather than teachers.

- ‘Mathematics makes me nervous.’
- ‘I am good at working out difficult mathematics problems.’
- ‘My teacher tells me I am good at mathematics.’
- ‘Mathematics is harder for me than any other subject.’
- ‘Mathematics makes me confused.’

The survey organisers have used pupils’ responses to these questions to create the Students Confidence in Mathematics scale, which is available within the public-use TIMSS database (variable ASBGSCM / BSBGSCM). Cronbach’s alpha for this scale within England is reported to be 0.87 for Year 5 pupils and 0.89 for Year 9 pupils. This scale is the secondary outcome within the analysis. For further details about how this scale has been constructed, see Martin, Mullis and Hooper (2016: chapter 15).

Sample size

As this project is a secondary analysis of data that has already been collected, the sample size should be considered fixed. It is widely recognised that ex-post power calculations should be avoided (Jiroutek and Turner, 2017) and hence are not presented here.

However, for readers interested in the minimum detectable effect size of the estimates presented, these can be calculated as approximately 2.8 times the reported standard errors.[7]

Selection mechanism

Why might teachers choose to group pupils within their class by attainment?

Little is currently known about why some teachers choose to teach children within same-attainment groups and others choose to teach within mixed-attainment groups (Francis, Taylor and Tereshchenko, 2019). The first research question hence provides new evidence on this issue in terms of observable child, teacher, and class characteristics in an exploratory analysis. A priori, two factors are thought to be key.

First, teachers choosing to group children by attainment is likely to be a reaction to the heterogeneity of their class. If it is very diverse, including children from a wide range of backgrounds and abilities, then some form of within-class attainment grouping may be a natural response. This then relates to other school policies, such as setting and streaming: classes are likely to be less diverse in those schools where some form of higher-level attainment grouping has already taken place. Likewise, it is also likely to be related to the diversity of a school’s intake: attainment grouping (whether within classes, between classes, or both) may be more appealing to teachers and school leaders when the student body is more diverse.

Second, within-class attainment grouping may be linked to individual teacher characteristics. For instance, more experienced teachers, or those with a specialisation in the subject they teach, may be more willing or confident to teach pupils in a certain way (such as using mixed-attainment groups). On the other hand, teachers may have been trained to use a certain approach during their initial training and have continued to use this approach throughout their career. In addition, it could be school policy (or the advice of a mentor) that results in pupils being grouped together in particular ways.

What determines the within-class attainment group to which children are assigned?

Within classes where children are grouped by attainment, what is likely to determine their allocation to the top, middle, or bottom group?

[7] See https://blogs.worldbank.org/impactevaluations/why-ex-post-power-using-estimated-effect-sizes-bad-ex-post-mde-not

In a perfectly meritocratic world, group allocation would be solely determined by two factors: (a) the child's own academic attainment and (b) the academic attainment of their class peers. Hence their *absolute* and *relative* academic attainment would be the two things that matter.

Yet, in reality, other factors are likely to determine group allocation. For instance, it may be that teachers' *perceptions* of a child's attainment (rather than their objectively measured attainment) is the more important factor. This could be influenced by their views of the child's motivation, ethnicity, or socio-economic background (Muijs and Dunne, 2010). Likewise, there could be other influences upon the group the child is assigned to, such as the 'pushiness' of parents or the preferences of a particular child (such as whether they would prefer to sit with their friends). Indeed, much research evidence (though mostly in the context of between-school or between-class tracking) has suggested that attainment group allocation is determined by factors outside of children's (measured) academic achievement, with those from lower socio-economic status and ethnic minority backgrounds being more likely to be allocated to lower sets (Connolly et al., 2019; Muijs and Dunne, 2010). Moreover, within-class attainment groups are unlikely to be fixed—their composition might well change over time.

Statistical analysis

Descriptive statistics

Recognising that little is currently known about the 'selection mechanisms' (why some teachers choose to teach in attainment groups while others do not) the analysis begins by looking at the characteristics of pupils exposed to within-class attainment grouping and those who are not (as well as documenting the characteristics of teachers who are most likely to organise their class in this way). These descriptive statistics illustrate the characteristics of pupils who are exposed to (a) same-attainment grouping, (b) mixed-attainment grouping, and (c) whole-class teaching within their mathematics classes. Cross-tabulations are also presented, documenting the association between mixed-attainment grouping, same-attainment grouping, and whole-class teaching in mathematics classes in England.

Primary analysis

The primary analysis is designed to address the following research question, with respect to Year 2, Year 5, and Year 9 pupils:

- Is within-class attainment grouping associated with higher levels of achievement?

Primary analysis of the Millennium Cohort Study (Year 2 pupils)

The primary analysis utilises an OLS regression model of the form:

$$A_{ij} = \alpha + \beta.G_{ij} + \partial.D_{ij} + \delta.P_{ij} + \gamma.S_{ij} + \Delta.T_{ij} + \tau.C_{ij} + \varepsilon_{ij} \quad (1)$$

Where:

i = Child i.

j = Teacher / class j.

A_{ij} = The achievement of child i in the age seven Progress in Mathematics test. This has been standardised (using the full age seven MCS sample) to mean zero and standard deviation one.

G_{ij} = A dummy variable capturing whether attainment grouping is used in the child's class (1) or not (0).

D_{ij} = A vector of children's demographic background characteristics, including gender, permanent family income, age in months, ethnicity, language spoken at home, and maternal and paternal education.

P_{ij} = Prior achievement of child i, as measured by performance in the MCS age five reading test, pattern construction scores, picture similarity scores, and children's Foundation Stage Profile scores (where available).[8]

S_{ij} = Whether the teacher reports that setting or streaming is used for the subject in the child's school, and the set to which the child has been assigned (bottom, middle, top).

T_{ij} = A vector of teacher characteristics. This includes their gender, years of teaching experience, qualifications, and years working at the school.

C_{ij} = A vector of class characteristics, as reported by the teacher. This includes class size, whether the class contains mixed year groups, percentage of children in class with SEN, percentage of children in class with EAL, number of days taught by a supply teacher, class time spent on literacy and numeracy per week, whether class gets regular support from a teaching assistant, special needs, or other teacher, and percentage of lesson time devoted to group work per week.

ε_{ij} = The error term. Standard errors are calculated having accounted for the stratification and clustering in the MCS survey design following the recommendations of the MCS documentation (see Ketende and Jones, 2011).

The parameter of interest from this model is β; this illustrates the difference in (standardised) Progress in Maths scores between children grouped by attainment within their class to those not grouped by attainment (conditional upon other variables included in the model). Multiple imputation using chained equations is used to account for missing covariate data with the age seven response weight (variable DOVWT2) applied. Standard errors are reported, adjusting for the stratification and clustering within the MCS survey design, as recommended in the technical documentation (see Ketende and Jones, 2011). All estimates are presented in terms of an effect size.

Trends in Mathematics and Science Study

The following OLS regression model is estimated, comparing the 'effect' of same-attainment grouping to the 'effect' of whole-class teaching:

$$A_{ijk} = \alpha + \beta.SA_{ij} + \gamma.WC_{ij} + \emptyset.P_i + \delta.Ch_i + \nabla.C_{ij} + \varphi.T_j + \omega.Set_k + \theta.Sch_k + \varepsilon_{ijk} \quad (2)$$

Where:

i = child i

j = teacher j

k = school k

A_{ijk} = The TIMSS mathematics plausible values, capturing children's achievement in this subject. These are standardised to mean zero and standard deviation one within England (using the full sample standard deviation). All model estimates are therefore presented in terms of effect sizes.

SA_{ij} = The variable capturing the frequency that same-attainment teaching is used within the child's mathematics class.

WC_{ij} = The variable capturing the frequency that whole-class teaching is used within the child's mathematics class.

P_i= Measures of children's prior achievement. For Year 5 pupils, this refers to their scores in their Key Stage 1 English and mathematics test. For Year 9 pupils, this refers to their Key Stage 2 English and mathematics test scores.

[8] Foundation Stage Profile scores are only available for children in England. They are therefore treated as missing data (and hence imputed) for children in other parts of the UK.

Ch_i = A vector capturing children's background characteristics. This includes gender, FSM status, home educational resources index—a measure encompassing books at home, home study supports, and parental education (where available), immigrant status, absence from school, and age in months.

T_j = A vector of teacher characteristics that may confound the relationship between the use of within-class attainment grouping and children's achievement. This includes key measures of observable teacher quality that have previously been identified in the academic literature, including (a) subject specialism in mathematics and (b) number of years teaching experience. It also includes characteristics of the class: (c) minutes per week spent teaching mathematics to the class, (d) average prior maths achievement levels of the class (based upon Key Stage scores), and (e) standard deviation of the prior maths achievement of the class.[9]

Set_k = A dummy variable measured at the school level for whether it is school policy to set/stream pupils into different mathematics classes.

Sch_k = A school-level control for headteacher reports of the percentage of pupils who come from advantaged/disadvantaged backgrounds.

The parameters of interest from this model are β and γ. The former (β) illustrates whether more frequent use of same-attainment grouping by teachers is associated with greater academic progress made in mathematics. The latter (γ) has a similar interpretation with respect to whole-class teaching. By comparing β to γ, one can establish which approach is most strongly associated with the progress children make in mathematics.

An analogous model is then estimated—including the same set of controls—to compare the impact of same-attainment grouping to mixed-attainment grouping:

$$A_{ijk} = \alpha + \beta.SA_{ij} + \gamma.MA_{ij} + \emptyset.P_i + \delta.Ch_i + \nabla.C_{ij} + \varphi.T_j + \omega.Set_k + \theta.Sch_k + \varepsilon_{ijk} \quad (3)$$

Where:

MA_{ij} = The variable capturing the frequency that mixed-attainment teaching is used within the child's mathematics class.

A comparison of the β and γ parameters now reveals whether same-attainment or mixed-attainment grouping is more strongly associated with mathematics achievement.

To account for the complex TIMSS survey design, the recommended practise of the survey organisers is followed. Specifically, the final TIMSS pupil and replication weights are applied via the Stata package 'pv' (Macdonald, 2008). This accounts for the hierarchical nature of the data, including the nesting of pupils within classes and schools, making the appropriate adjustment to the estimated standard errors. Likewise, following recommended practise in the use of plausible values, all models are estimated five times (that is, five versions of the model are estimated, with plausible value 1 used as the dependent variable in model 1, plausible value 2 used in model 2, and so-forth) with final parameter estimates and standard errors combined according to Rubin's multiple imputation rules (Rubin, 1987).

Multiple hypothesis testing

A Bonferroni correction is used to adjust for multiple hypothesis testing. As three independent datasets are used with three primary outcomes (mathematics achievement in each of the three datasets), a Bonferroni correction factor of three is applied (that is, the Bonferroni-adjusted critical p-value threshold is 0.05/3 = 0.0167).

The same analysis model is estimated as for the primary analysis but using logistic regression rather than ordinary least squares. This model is specified as:

$$E_{ij} = \alpha + \beta.G_{ij} + \partial.D_{ij} + \delta.P_{ij} + \gamma.S_{ij} + \Delta.T_{ij} + \tau.C_{ij} + \varepsilon_{ij} \quad (4)$$

Where:

[9] The average and standard deviation of prior achievement of the class are based upon Key Stage 1 maths scores (Year 5) and Key Stage 2 maths scores (Year 9).

E_{ij} = Whether the child enjoys doing number work a lot (1) or likes it a little or not at all (0).

All other variables included are as defined for equation (1).

The parameter of interest (β) illustrates whether children grouped by attainment within their class are more likely to enjoy number work than those not grouped by attainment within their class (conditional upon the other variables within the model). Results from this model are reported as log odds-ratios.

Moreover, the following additional secondary research question is answered using the MCS:

- Is being placed in a higher within-class attainment group associated with greater academic progress? (Year 2 pupils only.)

To answer this question, the sample is first restricted to children where within-class attainment grouping is used. An OLS regression model is then estimated, specified as:

$$A_{ij} = \alpha + \beta . G_{ij} + \partial . C_{ij} + \delta . P_{ij} + \gamma . S_{ij} + \Delta . T_{ij} + \tau . C_{ij} + \varepsilon_{ij} \quad (5)$$

Where:

G_{ij} = A dummy variable capturing the child's within-class attainment group; low, middle (reference), or high.

And all other variables are specified as in equation (1). The β estimates now capture whether children assigned to higher within-class attainment groups have higher levels of achievement than children assigned to lower within-class attainment groups—conditional upon the covariates (including the numerous measures of prior achievement) in the model.

This is supplemented using a child fixed-effects model capturing within-child, between-subject (English versus mathematics) attainment group allocation. To begin, the sample is restricted to children where within-class attainment grouping is used. These models exploit the fact that children may be allocated to different attainment groups in different subjects (for example, the middle group in English and the top group in mathematics). This between-subject variation potentially allows one to get closer to a 'causal' effect than regression/matching techniques alone. Specifically, such models implicitly control for all factors that are constant for a child (for example, innate attainment, socio-economic background, and school) with the focus now on children's *relative* performance in English versus mathematics (and how this relates to their attainment group allocation). The specification of this fixed-effects model is:

$$A_{ij} = \alpha + \beta . G_{ij} + \gamma . S_{ij} + \delta . P_{ij} + \Delta . T_{ij} + u_j + \varepsilon_i \quad (6)$$

Where:

i = subject i (mathematics or English).

j = child j.

A_{ij} = Achievement test score in subject i of child j.

G_{ij} = Within-class attainment group (low, middle, high) in subject i of child j.

S_{ij} = A vector of dummy variables capturing the set or stream (low, middle, high) in subject i of child j.

P_{ij} = Prior achievement of child j in subject i, as measured by scores in the Foundation Stage profile.[10]

T_{ij} = Amount of class time devoted to subject i in child j's class.

u_j = Child fixed-effect.

[10] Only foundation stage profile scores are included in this model. This is because the model now focuses upon *within-subject* variation and hence any controls must be subject specific. The foundation stage profile is the only direct source of prior achievement information in mathematics. Note also that this model is therefore only estimated for pupils in England.

ε_i = Random (within-child) error term.

The β parameter now captures how attainment group allocation in mathematics/English is related to relative performance in English/mathematics at age seven. The model also includes controls for factors that potentially vary within subject for each child such as set allocation and prior achievement in English and mathematics. Note that, when estimating these models, only the first imputed value for missing covariate data is used and the survey weight is not applied. The main implication of this is that the standard errors within this part of the analysis may be slightly underestimated due to not taking into account the uncertainty due to missing data.

Trends in Mathematics and Science Study

The secondary analysis of TIMSS addresses the research question:

- Is there an association between within-class attainment grouping and children's academic self-confidence (Year 5 and Year 9)?

See above for further details on how children's academic self-confidence is measured within TIMSS. The same analytic model is used as for the analysis of the primary outcome (maths achievement), but now using mathematics self-confidence as the dependent variable.

Missing data analysis

In Appendix A, the characteristics of pupils included in the final analytic samples are compared to those pupils excluded from the sample due to missing data. In the MCS, children excluded from the analytic sample are more likely to come from lower income backgrounds, have lower levels of prior achievement, and are less likely to be White. For Year 5 pupils, those excluded from the analysis are more likely to come from an immigrant background and to not speak English at home. They also study in slightly smaller class sizes and attend slightly more disadvantaged schools. For Year 9 pupils, those excluded due to missing data are more likely to come from immigrant backgrounds, have less experienced maths teachers, and are slightly more likely to be male. Otherwise, for both Year 5 and Year 9, pupils excluded from the TIMSS sample are broadly similar to those in the analytic sample in terms of observable characteristics.

Multiple imputation using chained equations is used to handle missing covariate data in the MCS analysis. Listwise deletion is used in the headline TIMSS analysis, with the robustness of results tested using multiple imputation as an alternative approach, as specified a priori in the project study plan.

Subgroup analyses

Millennium Cohort Study

Three sources of possible heterogeneity are investigated: differences by socio-economic background, differences by prior achievement, and differences by whether other forms of attainment grouping (such as setting or streaming) are used within the school. These subgroups are defined as follows:

- *Family background*. Permanent family income is the primary measure of family background used in the MCS analysis. This is defined as follows. First, an average is taken of the income measures collected in the first five MCS waves (using, for each child, all waves where data is available). This permanent income measure is then divided into thirds (tertiles) to create low, average, and high income groups.
- *Prior achievement*. Children's scores on the age five MCS cognitive tests (naming vocabulary, pattern similarities, and picture similarities) are used to define children's prior achievement. To begin, scores on each of these tests are standardised to mean 0 and standard deviation 1. The average is then calculated for each child across these three test scores. This measure is then divided into tertiles to create low (bottom third), average (middle third), and high (top third) prior achievement groups.
- *Setting or streaming in the school*. As part of the MCS, teachers were asked about whether setting or streaming were used in their school and for what subjects. The survey organisers have derived a set of variables to indicate whether children are set or streamed in mathematics (variable strnum).

This variable is used to identify subgroups of children who have been grouped into different classes based upon their prior attainment or achievement.

To explore subgroup effects, the same model as presented in (1) is re-estimated for (a) high-income pupils, (b) low-income pupils, (c) children with low prior achievement, (d) children with high prior achievement, (e) children who have been set or streamed, and (f) children who have not been set or streamed.

Trends in Mathematics and Science Study

Separate results are presented for the following subgroups.

- Children from a socio-economically disadvantaged versus non-disadvantaged background. This is operationalised in TIMSS as the performance of children who are FSM-eligible (disadvantaged pupils) and those who are not FSM-eligible (non-disadvantaged pupils). Note that, throughout the analysis, the FSM group is defined as any child who has been eligible for free school meals over the previous six years (that is, EverFSM6).
- Children are divided into low, average, and high prior achievement groups based on the tertiles of their Key Stage 1 (Year 5) or their Key Stage 2 (Year 9) mathematics scores.
- The TIMSS Year 5 sample is divided into schools where setting or streaming is used and those where it is not. This enables investigation of the association between within-class attainment grouping and children's achievement in settings where they have already been separated by attainment into different classes (or not). Note this analysis is not conducted for the Year 9 TIMSS sample as almost all secondary schools in England set or stream pupils in mathematics.

The main analysis model presented above is hence estimated for the following subgroups: (a) FSM, (b) non-FSM, (c) low prior achievement, (d) high prior achievement, (e) setting used in school, and (f) setting not used in school.

Additional analyses and robustness checks

Millennium Cohort Study

Within certain elements of the analysis, a lack of 'common support' may be a concern. For instance, when investigating whether attainment group allocation matters, it is possible that there are no sufficiently comparable children in the top and bottom groups. The main advantage of propensity score matching (PSM) over regression is that common support can be enforced on the analytic sample thus ensuring that there are comparable children in the 'treatment' (for example, top attainment group) and control (for example, bottom attainment group) groups. PSM is therefore used as an alternative to test the robustness of our regression results.

The PSM models match children using one-to-one nearest-neighbour matching with the caliper set to 0.05. The first-stage selection model includes the same covariates as outlined for the regression models in equation (1). Estimates are based upon the first imputed dataset.[11]

Finally, due to a delay with the fieldwork, the MCS questionnaire was sent to teachers relatively late in the academic year. Consequently, some teachers completed the questionnaire after the child had progressed from Year 2 to Year 3, which could introduce problems with recall. The primary analysis model presented in equation (1) is therefore re-estimated using the subsample of teachers who completed the questionnaire before the start of the new academic year (before September 2008).

Trends in Mathematics and Science Study

To test the robustness of the results, the headline analysis models are re-estimated including some additional controls. The motivation for not including these controls in the headline model specification is due to some concerns over overfitting the data (given the limited number of teachers in TIMSS).

---

[11] The study protocol suggested that listwise deletion or 'missing' dummy flags would be used instead. However, the amount of missing data on covariates meant that listwise deletion was not feasible. It was therefore decided that the PSM models would be estimated using the first imputed variables of the covariates.

Specifically, the following additional variables are added to the model:

- class size;
- teachers' confidence in teaching mathematics;
- school shortages in mathematics; and
- language spoken at home by the pupil.

The second robustness test uses multiple imputation to account for the limited amount of missing covariate data, which mostly stems from non-linkage to the NPD or teacher non-response to the background questionnaire. This is implemented via multiple imputation using chained equations in Stata with the random seed set to a value of 5000 producing five imputed datasets. The headline analysis is re-run using this imputed data with the first plausible value in mathematics as the outcome and a Huber-White adjustment made to the estimated standard errors at the school level.

Estimation of effect sizes

Effect sizes are presented throughout this report. These have been calculated as follows. In the MCS and TIMSS datasets, outcomes are standardised to mean zero and standard deviation one. This has been done by subtracting the (weighted) mean and dividing the (weighted) standard deviation from each pupil's mathematics scores. These standardised variables are then used as the outcomes within the analysis models, as specified above. The parameter estimates from the models hence refer to associations in terms of an effect size.

Timeline

Table 3: Timeline

Dates	Activity
October 2019	Study plan written.
12 November 2019	Advisory board meeting.
6 December 2019	TIMSS-NPD access form submitted to DfE.
17 December 2019	Ethical approval granted.
21 January 2020	EEF grants committee approved project.
February 2020	Analysis of MCS and TIMSS (public-use) data files.
March 2020	Draft report (using MCS and public-use TIMSS data) submitted to EEF.
April 2020	EEF review draft report.
May 2020	Draft report revised following EEF comments.
September 2020	Revised statistical analysis using TIMSS-NPD linked data.
October 2020	Report revised based upon linked analysis.

Impact evaluation results

Descriptive statistics

Millennium Cohort Study

Table 4 compares the characteristics of pupils (and their teachers) in situations where within-class attainment grouping is used in their mathematics and English lessons. There are few differences by demographic characteristics and levels of prior achievement: Year 2 pupils are just as likely to be taught in (within-class) attainment groups regardless of their gender, ethnicity, family background, and prior levels of academic achievement. This holds true for both English and mathematics. Likewise, there is little clear association between the characteristics of Year 2 class teachers and the use of within-class attainment grouping: the distribution of teacher experience and qualifications is similar across the two groups.

Indeed, the only clear difference that emerges in Table 4 is with respect to setting and streaming. In particular, for those children not placed into attainment groups within their mathematics classes, 58% are taught in (between-class) attainment sets and 20% in (between-class) attainment streams. This compares to 29% (setting) and 15% (streaming) for those children where within-class attainment grouping is used. A similar pattern emerges for English. In other words, within-class attainment grouping is more common in schools where setting or streaming is not used.

Table 4: Characteristics of Year 2 pupils and teachers where within-class attainment grouping is used

	Mathematics		English	
Within-class attainment grouping used?	Yes	No	Yes	No
Child background				
Average family income (equivalised £ per week)	363	374	365	358
% father NVQ level 4/5	28%	31%	29%	30%
% mother NVQ level 4/5	32%	35%	32%	34%
% White	89%	85%	89%	81%
% Male	50%	50%	50%	50%
Child's prior achievement				
Total FSP score (England only)	0.14	0.09	0.15	0.01
Bracken school readiness score	0.06	0.05	0.07	0.00
Age 5 pattern construction score	0.08	0.09	0.09	0.02
Age 5 picture similarities score	0.00	0.08	0.01	0.06
Age 5 vocabulary score	0.13	0.11	0.14	0.04
Teacher characteristics				
Years of teaching experience	13.5	13.6	13.5	13.7
Holds PGCE	34%	31%	34%	29%
Holds other postgraduate degree	5%	9%	6%	6%
Class characteristics				
Children set	29%	58%	25%	47%
Children streamed	15%	20%	16%	18%
Class size	26	24	26	24
Mixed year groups	26%	29%	26%	30%
% EAL	9%	10%	9%	11%
% of class time spent working in groups	31%	29%	31%	29%
Teaching assistant in classroom	86%	82%	86%	82%
Approximate number of children	**6,345**	**1,295**	**6,527**	**1,061**

Note: Number of observations approximate due to differing amounts of missing data across variables.

Amongst those pupils where within-class attainment grouping is used, Table 5 compares the characteristics of pupils allocated to the top, middle, and bottom groups. Children allocated to lower within-class attainment groups tend to come from lower-income backgrounds, have lower levels of prior achievement, and (in the case of English) are more likely to be male. In other words, lower-achieving and less advantaged children are more likely to find themselves in the bottom within-class attainment groups.

Table 5: Characteristics of Year 2 pupils allocated to the top, middle, and bottom within-class attainment groups

	Mathematics			English		
Within-class attainment group?	Low	Middle	High	Low	Middle	High
Child background						
Average family income	303	349	403	300	356	410
% father hold degree	17%	25%	37%	18%	25%	38%
% mother hold degree	20%	30%	40%	21%	31%	40%
% White	88%	88%	90%	89%	90%	89%
% Male	53%	47%	52%	64%	50%	43%
Child's prior achievement						
Total FSP score (England only)	-0.78	0.08	0.63	-0.77	0.10	0.69
Bracken school readiness score	-0.55	-0.06	0.43	-0.56	-0.03	0.48
Age 5 pattern construction score	-0.53	0.03	0.41	-0.43	0.07	0.39
Age 5 picture similarities score	-0.40	-0.07	0.25	-0.32	-0.07	0.25
Age 5 vocabulary score	-0.36	0.05	0.42	-0.33	0.08	0.45
Approximate number of children	**1,094**	**2,457**	**2,794**	**1,275**	**2,524**	**2,728**

Note: Number of observations approximate due to differing amounts of missing data across variables.

Trends in Mathematics and Science Study Year 5

Table **6** begins by illustrating the frequency that Year 5 pupils in England are exposed to different teaching approaches during their mathematics lessons. Just over a third of pupils experience whole-class teaching in almost every lesson, around 40% in half of lessons, and a quarter in some or no lessons. The percentages for same-attainment grouping are similar. On the other hand, within-class mixed-attainment grouping occurs less frequently, with just over half of Year 5 pupils in England being subject to this approach in some or no lessons. Table **6** hence reveals how within-class mixed-attainment grouping occurs less frequently in Year 5 classrooms in England than either same-attainment grouping or whole-class teaching.

Table 6: The frequency that Year 5 pupils are exposed to mixed-attainment, same-attainment, and whole-class teaching approaches in mathematics

	Same-attainment grouping	Mixed-attainment grouping	Whole-class teaching
Every/almost every lesson	42%	21%	37%
About half of lessons	36%	23%	41%
No/some lessons	22%	56%	23%
Observations	**2,856**	**2,856**	**2,856**

Note: Figures refer to column percentages.

Table **7** turns to the association between the use of mixed-attainment and same-attainment grouping: are these practises mutually exclusive, or do teachers use a combination of the two? There is, as expected, an inverse relationship: Year 5 pupils who are exposed to mixed-attainment grouping in their mathematics class are less likely to be exposed to same-attainment grouping. For instance, 65% of pupils who are rarely placed in mixed-attainment groups are placed into same-attainment groups in almost every maths lesson. This is compared to 21% of Year 5 pupils who reportedly experience both approaches in almost all of their mathematics classes.

Table 7: The association between mixed and same-attainment grouping of pupils in mathematics teaching—Year 5 pupils

Same-attainment ↓ / mixed-attainment →	Every/almost every lesson	About half of lessons	No/some lessons
Every/almost every lesson	21%	9%	65%
About half of lessons	32%	91%	15%

No/some lessons	47%	0%	20%
Observations	**624**	**631**	**1601**

Note: Figures refer to column percentages.

Table **8** presents analogous results for the association between the frequency of whole-class teaching and within-class attainment grouping of pupils. It is striking that most Year 5 pupils (61%) who receive whole-class teaching in nearly every maths lesson also work in attainment groups with equal frequency. In contrast, Year 5 pupils in classes where whole-class teaching is rarely used are most likely to work in same-attainment groups around half their lessons (43%). It again seems that primary teachers in England use a mix of approaches to mathematics instruction.

Table 8: The association between whole-class teaching and same-attainment grouping of pupils in mathematics teaching—Year 5 pupils

Same-attainment ↓ / whole-class →	Every/almost every lesson	About half of lessons	No/some lessons
Every/almost every lesson	61%	32%	32%
About half of lessons	24%	43%	43%
No/some lessons	15%	26%	24%
Observations	**1,129**	**1,105**	**622**

Notes: Figures refer to column percentages.

Finally, Table **9** considers a similar trade-off between whole-class teaching and mixed-attainment grouping. Again, those Year 5 pupils who regularly experience whole-class teaching are the most likely to also work in mixed-attainment groups during their mathematics classes (31%). This reduces to 5% for those who receive whole-class teaching in only some or no lessons.

Table 9: The association between whole-class teaching and mixed-attainment grouping of pupils in mathematics teaching—Year 5 pupils

Mixed-attainment ↓ / whole-class →	Every / almost every lesson	About half of lessons	No / some lessons
Every / almost every lesson	31%	22%	5%
About half of lessons	14%	26%	32%
No/some lessons	56%	51%	63%
Observations	**1,129**	**1,105**	**622**

Notes: Figures refer to column percentages.

Table **10** turns to the association between the frequency each of the three different teaching approaches are used and the characteristics of Year 5 pupils. This table therefore reveals whether certain types of pupils (and certain types of teachers) are more likely to experience (or, in the case of teachers, deliver) particular approaches. There is little clear association between pupil characteristics and the frequency that their mathematics teacher uses different approaches. These characteristics (gender, immigrant status, English spoken regularly at home, and frequency of absence from school) are hence unlikely to confound the results. In contrast, Year 5 pupils who are rarely taught maths in same-attainment groups are more likely to have a teacher with a background in mathematics (22%) than pupils who are taught in same-attainment groups in most maths lessons. There also seems to be differences in the approach used when teaching experience is taken into account. Year 5 pupils regularly taught in same-attainment groups tend to have less experienced teachers (eight years of experience on average) than their peers who are taught in same-attainment groups infrequently (11 years of experience on average). Less experienced teachers are also less likely to teach mixed-attainment groups.

Another key point of difference includes average class sizes being lower when same-attainment grouping is used infrequently. Similarly, the average amount of mathematics instruction received per week is higher when mixed-attainment grouping occurs regularly compared to infrequently (338 versus 290 minutes), with the opposite holding true for same-attainment grouping (329 minutes per week for pupils taught in same-attainment groups in none/some lessons compared to 290 minutes where same-attainment grouping is used in almost every lesson).

Finally, schools with a high-proportion of disadvantaged pupils seem more likely to use whole-class teaching. Amongst teachers who use whole-class teaching in almost every maths lesson, 33% report most pupils in their school come from a disadvantaged background. This compares to 17% of those where whole-class teaching is used infrequently.

Table 10: Association between background characteristics and mathematics teaching approach—TIMSS Year 5 sample

	Same-attainment groups			Mixed-attainment groups			Whole-class teaching		
	Never/ some	Half	Almost every	Never/ some	Half	Almost every	Never/ some	Half	Almost every
Child characteristics									
% male	47%	46%	50%	48%	48%	47%	47%	47%	50%
% FSM	30%	27%	34%	33%	29%	28%	29%	25%	38%
% immigrant	10%	9%	9%	9%	11%	8%	8%	10%	10%
% English not spoken at home	17%	16%	14%	16%	16%	15%	12%	15%	18%
% regularly attends school	87%	92%	90%	89%	92%	90%	92%	91%	88%
Prior achievement									
Key Stage 1 maths score	16.1	16.6	16.2	16.3	16.4	16.4	16.6	16.4	16.2
Key Stage 1 reading score	15.7	16.1	15.5	15.6	15.8	16.0	16.0	15.8	15.5
Teacher characteristics									
% subject specialism in maths	22%	8%	9%	9%	9%	20%	8%	14%	10%
Years of teaching experience	10.7	13.0	8.0	9.4	11.9	11.1	10.9	11.1	9.1
Class characteristics									
Class size	26.9	27.7	28.2	28.0	26.5	28.4	27.5	27.9	27.7
% students with language issues	11%	8%	7%	7%	6%	13%	4%	9%	10%
Minutes per week teaching class	329	304	290	290	302	338	305	302	303
Average KS1 maths scores of class (averages of class means)	16.1	16.6	16.2	16.3	16.3	16.4	16.6	16.3	16.1
Standard deviation of class KS1 maths scores	3.2	3.0	2.8	2.9	3.1	3.0	3.2	3.0	2.9
School characteristics									
Pupils set for mathematics	59%	58%	56%	54%	41%	81%	45%	40%	84%
0–10% of pupils from poor backgrounds	20%	25%	13%	15%	25%	22%	23%	24%	10%
11–25% of pupils from poor backgrounds	32%	32%	33%	32%	36%	31%	31%	42%	24%
26–50% of pupils from poor backgrounds	27%	16%	20%	24%	7%	25%	24%	17%	22%
51–100% of pupils from poor backgrounds	16%	12%	26%	22%	15%	14%	17%	7%	33%
Resource shortages affecting maths instruction scale	0.24	-0.25	-0.02	-0.07	-0.42	0.41	-0.28	-0.10	0.18
Observations	**2856**			**2856**			**2856**		

Note: The 'resource shortages affecting maths instruction scale' is based upon headteacher responses about shortages or the inadequacy of 13 items. These included a shortage of teachers with a specialisation in maths, too few calculators, and a lack of access to a range of general school resources (such as buildings or instructional materials). See Martin, Mullis and Hooper (2016) for further details.

Trends in Mathematics and Science Study Year 9

Table **11** reveals that whole-class teaching is the most common approach to mathematics instruction in Year 9 classes. In total, 38% of Year 9 pupils receive whole-class teaching in almost every mathematics lesson, with a further 37% experiencing this in half their maths lessons. On the other hand, within-class mixed-attainment grouping of pupils is comparatively rare; just 8% of Year 9s are taught in mixed-attainment groups within their mathematics class in almost every lesson, with 70% suggesting that mixed-attainment grouping occurs in only 'some' or 'no' maths lessons. Teaching Year 9 pupils in same-attainment groups is somewhat more common, though still almost half of Year 9 pupils only experience this teaching approach in some/none of their maths lessons.

Table 11: The frequency that Year 9 pupils are exposed to mixed-attainment, same-attainment, and whole-class teaching approaches in mathematics

	Same-attainment grouping	Mixed-attainment grouping	Whole-class teaching
Every/almost every lesson	28%	8%	38%
About half of lessons	20%	22%	37%
No/some lessons	52%	70%	24%
Observations	**3,288**	**3,288**	**3,288**

Note: Figures refer to column percentages.

Table **12** turns to the association between mixed-attainment grouping and same-attainment grouping in Year 9 mathematics lessons. Amongst the majority of teachers who rarely use mixed-attainment grouping, 38% use same-attainment grouping in almost every mathematics lesson. Yet 57% of Year 9 pupils who are rarely taught in mixed-attainment groups are also rarely taught in same-attainment groups. There is also some evidence that some Year 9 pupils in England receive a mixed approach, with 73% of those taught in mixed-attainment groups for half their maths lessons taught in same-attainment groups in the other half.

Table 12: The association between mixed and same-attainment grouping of pupils in mathematics teaching—Year 9 pupils

Same-attainment ↓ / mixed-attainment →	Every/almost every lesson	About half of lessons	No/some lessons
Every/almost every lesson	13%	3%	38%
About half of lessons	6%	73%	5%
No/some lessons	81%	25%	57%
Observations	**305**	**749**	**2234**

Note: Figures refer to column percentages.

The equivalent trade-off between whole-class teaching and same-attainment grouping is presented in Table **13**. Interestingly, the distribution across the three columns is reasonably similar. For instance, around a third of Year 9 pupils experience same-attainment grouping in almost every lesson, regardless of how frequently whole-class teaching is used. Likewise, around half of Year 9s are only taught in same-attainment groups in no/some lessons, with little variation by whether whole-class teaching is used in every lesson (51%), about half of lessons (52%) or no/some lessons (54%). Consequently, in Year 9 mathematics classes, the use of whole-group teaching and same-attainment grouping appear to be somewhat independent of one another, with some teachers sticking exclusively to one approach and others mixing the two.

Table 13: The association between whole-class teaching and same-attainment grouping of pupils in mathematics teaching—Year 9 pupils

Same-attainment ↓ / whole-class →	Every / almost every lesson	About half of lessons	No / some lessons
Every / almost every lesson	30%	23%	33%
About half of lessons	19%	26%	13%
No/some lessons	51%	52%	54%
Observations	**1,253**	**1,270**	**765**

Notes: Figures refer to column percentages.

A similar result emerges for the association between whole-class teaching and the use of mixed-attainment grouping in Year 9 mathematics classes: the distribution across the three columns is again reasonably similar. Although there is some suggestion that those Year 9 pupils who regularly receive whole-class teaching are slightly less likely to be taught in mixed-attainment groups (in their class), the link between the two is relatively weak. The other point that Table **14** helps to reiterate is the infrequent use of mixed-attainment groups to teach Year 9 pupils maths.

Table 14: The association between whole-class teaching and mixed-attainment grouping of pupils in mathematics teaching—Year 9 pupils

Mixed-attainment ↓ / whole-class →	Every/almost every lesson	About half of lessons	No/some lessons
Every/almost every lesson	4%	10%	12%
About half of lessons	20%	24%	22%
No/some lessons	77%	66%	66%
Observations	**1,253**	**1,270**	**765**

Note: Figures refer to column percentages.

Table 15 turns to the characteristics of pupils who receive (and of the teachers/schools who deliver) the three different teaching approaches. As for the results for Year 5 pupils, there is little evidence of substantial variation by pupil characteristics. There are, however, sizeable differences in terms of prior achievement. In particular, Year 9 pupils taught in same-attainment groups in almost every lesson tend to have higher Key Stage 2 test scores than their peers who are taught in same-attainment groups less frequently. On the other hand, Year 9 pupils who are frequently taught in mixed-attainment groups tend to have lower Key Stage 2 scores than their peers who are rarely taught in mixed-attainment classes.

It is also striking how the average number of years' teaching experience is around four years higher amongst those that use mixed-attainment grouping rarely (11.3 years) compared to those who use it in half (8.2 years) or more (7.4 years) lessons. Likewise, it appears that less experienced teachers may be less likely to use whole-class approaches to teaching mathematics to Year 9 pupils.

Although there is little obvious variation between the frequency each approach is used and class size, there do seem to be differences according to teaching time. In particular, the average contact time is lower amongst those Year 9 pupils who are regularly taught in same-attainment groups (179 minutes per week) compared to those where same-attainment grouping is infrequent (203 minutes per week). Finally, in contrast to the results for Year 5 pupils, there is no evidence that schools with a greater proportion of disadvantaged children are more likely to favour a particular approach.

Table 15: Association between background characteristics and mathematics teaching approach—TIMSS Year 9 sample

	Same-attainment groups			Mixed-attainment groups			Whole-class teaching		
	Never / some	Half	Almost every	Never / some	Half	Almost every	Never / some	Half	Almost every
Child characteristics									
% male	46%	48%	53%	49%	49%	43%	45%	48%	51%
% FSM	26%	28%	24%	26%	26%	28%	25%	27%	26%
% immigrant	7%	8%	6%	6%	7%	11%	8%	8%	5%
% English not spoken at home	3%	5%	3%	3%	4%	5%	4%	4%	3%
% regularly attend school	92%	89%	93%	92%	91%	93%	91%	92%	92%
Prior achievement									
Key Stage 2 maths score	68.7	68.1	75.7	71.3	69.5	66.9	71.7	71.3	69.2
Key Stage 2 reading score	31.0	30.6	33.6	31.9	31.2	30.5	32.0	31.9	31.2
Teacher characteristics									
% Subject specialism in maths	78%	77%	87%	84%	72%	70%	81%	78%	82%
Years of teaching experience	10.6	9.0	10.7	11.3	8.2	7.4	8.4	10.2	11.6
Class characteristics									
Class size	26.4	26.4	26.9	26.3	26.9	27.3	26.6	26.1	26.8
% students with language issues	2%	7%	3%	2%	5%	7%	2%	3%	3%
Minutes per week teaching class	203	217	179	193	214	203	185	210	196
Average KS2 maths scores of class (averages of class means)	68.6	68.2	75.7	71.2	69.6	67.1	71.7	71.3	69.0
Standard deviation of class KS2 maths scores	10.0	11.0	9.0	9.6	10.5	11.2	9.2	10.3	10.1
School characteristics									
0-10% of pupils from poor backgrounds	15%	19%	20%	14%	22%	27%	20%	15%	17%
11-25% of pupils from poor backgrounds	35%	38%	36%	36%	36%	36%	29%	40%	36%
26-50% of pupils from poor backgrounds	21%	25%	16%	22%	19%	14%	21%	19%	22%
51-100% of pupils from poor backgrounds	12%	7%	10%	9%	12%	19%	12%	13%	7%
Resource shortages affecting maths instruction scale	0.09	0.08	-0.30	-0.05	-0.01	0.22	-0.32	0.12	0.04
Observations		**3288**			**3288**			**3288**	

Note: The 'resource shortages affecting maths instruction scale' is based upon headteacher responses about shortages or the inadequacy of 13 items. These included a shortage of teachers with a specialisation in maths, too few calculators, and a lack of access to a range of general school resources (such as buildings or instructional materials). See Martin, Mullis and Hooper (2016) for further details.

Outcomes and analysis

Millennium Cohort Study

Table **16** presents estimates of the association between within-class attainment grouping and the mathematics achievement of Year 2 pupils. These results are based upon the primary analysis regression model, as specified in the methodology section above. The estimated effect size is small (0.04, equivalent to zero months' additional progress using the EEF's conversion metric) and not statistically significant at conventional thresholds. In other words, there is no evidence that the use of within-class attainment grouping within Year 2 mathematics lessons leads to higher (or lower) levels of mathematics achievement (on average).

Table 16: The association between within-class attainment grouping and mathematics achievement amongst Year 2 pupils (primary outcome)

	Mathematics scores		
	Effect	SE	N
Within-class attainment grouping used?			
No (reference group)			
Yes	0.041	0.030	7899

Note: * indicates statistical significance at the 5% level after a Bonferroni correction factor of three has been applied (critical t-value of 2.394).

Equivalent results for the secondary outcomes are presented in Table **17**. This also includes an exploratory analysis of reading test scores (which was not pre-specified as part of the project protocol). A similar pattern emerges. There is no evidence that teaching within attainment groups in Year 2 lessons is any more (or less) likely to lead to children enjoying number work; the association is small (odds ratio = 1.04) and not statistically significant. Similarly, average reading scores do not seem to vary depending upon whether within-class attainment grouping is used or not (effect size = -0.04).

Table 17: The association between within-class attainment grouping, enjoyment of number work and reading achievement amongst Year 2 pupils (secondary outcomes)

	Enjoy number work			Reading scores		
	Logit	SE	n	Effect	SE	n
Within-class attainment grouping used?						
No (Reference group)						
Yes	0.040	0.074	7489	-0.043	0.033	7,742

Notes: * indicates statistical significance at the five% level after a Bonferroni correction factor of three has been applied (critical t-value of 2.394).

Separate results by subgroup for the primary outcome (mathematics achievement) are presented in Table **18** with analogous estimates for the secondary outcomes (enjoyment of number work and reading scores) presented in Appendix B. There are two subgroups where a modest, positive relationship is found. The first is for pupils from low-income backgrounds: the use of within-class attainment grouping is associated with a 0.14 standard deviation increase in their mathematics achievement (equivalent to two months' additional progress using the EEF conversion metric). However, the same finding does not emerge for the secondary outcomes (enjoyment of number work and reading test scores) reported in Appendix B, indicating that appropriate caution should be exercised when interpreting this result. Similarly, children with high levels of prior achievement seem to benefit, on average, from the practise of attainment-grouping in mathematics (effect size 0.10), though again the same pattern does not emerge for either of the secondary outcomes.

Table 18: The association between within-class attainment grouping and mathematics achievement amongst Year 2 pupils—sub-group analyses

	Mathematics		
	Effect	SE	n
Family-income			
Low-income (bottom third)	0.143*	0.058	2622
High-income (top third)	0.012	0.042	2625
Prior achievement			
Low prior achievement (bottom third)	-0.013	0.061	2523
High prior achievement (top third)	0.100	0.045	2536
Setting/streaming used for subject			
No	-0.020	0.048	4493
Yes	0.063	0.038	2830

Note: * indicates statistical significance at the 5% level after a Bonferroni correction factor of three has been applied (critical t-value of 2.394).

Robustness tests for these results are presented in Appendix C. The first set of robustness tests illustrate how results change if the analysis sample is first restricted to the subsample of teachers who completed the questionnaire before September 2008 (to minimise issues with recall bias). The second set of robustness tests uses propensity score matching rather than regression to control for potential confounding. Both lead to similar substantive conclusions: there is no clear evidence that the use of within-class attainment grouping is associated with higher (or lower) levels of achievement in mathematics.

Table **19** turns to differences in mathematics achievement between children allocated to the top and bottom within-class attainment groups compared to those allocated to the middle group. These estimates are based upon the primary analysis model specification as outlined in the methodology section above. Children allocated to the lowest within-class attainment group have lower levels of mathematics achievement than those allocated to the middle group (effect size difference of 0.41) who, in turn, have lower levels of mathematics achievement than their peers in the top group (effect size 0.34). Consequently—conditional upon prior achievement and other demographic background as well as teacher and class characteristics—children allocated to higher within-class attainment groups achieve higher mathematics test scores than those allocated to lower groups.

Table 19: The association between within-class attainment group a Year 2 child is allocated to and their mathematics achievement—OLS regression estimates

	Mathematics		
	Effect	SE	N
Within-class attainment group			
Low	-0.408*	0.037	6572
Middle (reference group)	-	-	
High	0.344*	0.029	

Note: * indicates statistical significance at the 5% level after a Bonferroni correction factor of three has been applied (critical t-value of 2.394).

Similar results emerge for the secondary outcomes (Table **20**). Year 2 pupils in the lowest within-class attainment group achieve lower reading test scores than those in the middle group (effect size 0.62) who themselves achieve lower scores than those in the highest attainment group (effect size 0.54). For enjoyment of number work, there is no difference between the lowest and middle attainment groups (odds ratio = 0.97). However, those Year 2 pupils allocated to the top within-class attainment group are more likely to enjoy number work than those in the middle or lowest groups (odds ratio = 1.51).

Table 20: The association between within-class attainment group a Year 2 child is allocated to and their (a) enjoyment of number work and (b) reading achievement

	Enjoy number work			Reading scores		
	Logit	SE	N	Effect	SE	n
Within-class attainment group						
Low	-0.032	0.092	6232	-0.615*	0.033	6692
Middle (reference group)	-			-		
High	0.414*	0.074		0.535*	0.025	

Note: * indicates statistical significance at the 5% level after a Bonferroni correction factor of three has been applied (critical t-value of 2.394).

To conclude the analysis of the MCS data, Table **21** presents the fixed-effect estimates. These compare relative performance in reading and mathematics amongst pupils assigned to different attainment groups (e.g. low versus middle) across the two subjects. The results are consistent with those presented above. There continues to be evidence that Year 2 children assigned to the lowest within-class attainment group achieve lower test scores than those assigned to the middle and the top groups. There is hence robust evidence that within-class group allocation is related to subsequent achievement.

Table 21: The association between within-class attainment group a Year 2 child is allocated to and their mathematics achievement. Child fixed-effects estimates.

	Effect	SE	n
Within-class attainment group			
Low	-0.364*	0.082	382
Middle (reference group)	-		
High	0.420*	0.064	

Notes: * indicates statistical significance at the five% level after a Bonferroni correction factor of three has been applied (critical t-value of 2.394).

Trends in Mathematics and Science Study Year 5

Results for the primary (mathematics achievement) and secondary (mathematics self-confidence) outcomes can be found in Table **22**. Specifically, this illustrates how these outcomes are related to same-attainment grouping and whole-class teaching. Overall, there is no evidence that more frequently using either approach is associated with higher levels of mathematics achievement of Year 5 pupils; the effect sizes associated with the same-attainment grouping and whole-class teaching variables are always small (below 0.1 standard deviations) and never statistically significant. There is hence no evidence that more frequently using whole-class teaching instead of same-attainment grouping (or vice-versa) is associated with higher levels of mathematics achievement.

A similar finding emerges with respect to the secondary outcome, self-confidence in mathematics. The effect size estimates are typically small and not statistically significant. The central finding of Table **22** is thus that it provides little clear evidence that more frequent use of either same-attainment grouping or whole-class teaching is associated with stronger mathematics outcomes.

Table 22: The association between same-attainment grouping, whole-class teaching, and mathematics achievement amongst Year 5 pupils

	Maths achievement			Self-confidence		
	Effect	SE	N	Effect	SE	n
Teaching same-attainment groups						
No/some lessons (reference)						
Half of lessons	0.011	0.087	2856	0.003	0.048	2856
Almost every lesson	0.033	0.096	2856	0.028	0.041	2856
Whole-class teaching						
No/some lessons (reference)						
Half of lessons	-0.053	0.079	2856	0.071	0.042	2856
Almost every lesson	0.036	0.087	2856	0.071	0.043	2856

Note: Estimates presented in terms of effect sizes; SE refers to the standard error.

Equivalent results comparing the association between mathematics achievement, same-attainment grouping, and mixed-attainment grouping for Year 5 pupils are presented in Table **23**. A similar substantive conclusion holds: there is no clear evidence that using one approach more frequently than the other is linked to higher levels of mathematics achievement or self-confidence. With respect to the former, the estimated associations in terms of effect sizes are small with no evidence of a 'dose-response' relationship. For instance, when looking at maths achievement as the outcome, all the estimated effect sizes are below 0.05 in terms of absolute magnitude. Moreover, the estimated standard errors are quite large, suggesting that there is quite sizeable uncertainty in the result. For the secondary outcome—mathematics self-confidence—the estimated effect sizes are also small (consistently below 0.1 standard deviations). Overall, Table **23** does little to suggest that using one approach more frequently than the other is associated with Year 5 pupils' achievement in mathematics.

Table 23: The association between same-attainment grouping, mixed-attainment grouping, and mathematics achievement amongst Year 5 pupils

	Maths achievement			Self-confidence		
	Effect	SE	N	Effect	SE	n
Teaching same-attainment groups						
No/some lessons (reference)						
Half of lessons	-0.022	0.089	2856	0.046	0.056	2856
Almost every lesson	0.037	0.102	2856	0.036	0.044	2856
Teaching mixed-attainment groups						
No/some lessons (reference)						
Half of lessons	0.036	0.091	2856	-0.074	0.040	2856
Almost every lesson	-0.035	0.106	2856	0.002	0.042	2856

Note: Estimates presented in terms of effect sizes; SE refers to the standard error.

Separate results by subgroup for the primary outcomes are presented in Table **24** (comparing same-attainment grouping to whole-class teaching) and Table **25** (comparing same-attainment to mixed-attainment grouping). Analogous results for the secondary outcome (mathematics self-confidence) are reported in Appendix B. For each subgroup of interest, there is little evidence that any one of the approaches is superior to the others. Most effect sizes are small and/or the estimates fail to reach statistical significance at conventional thresholds (due, in part, to the large standard errors). There is, for instance, no evidence that Year 5 FSM pupils achieve greater levels of progress in mathematics if whole-class teaching is used instead of within-class same-attainment grouping, or if within-class mixed-attainment grouping is used instead of within-class same-attainment grouping. The same hold true with respect to pupils with lower-levels of prior achievement and for pupils who are set or streamed into different classes.

Table 24: The association between same-attainment grouping, whole-class teaching, and mathematics achievement—Year 5 pupils, subgroup estimates

	FSM		Not FSM		Low achiever		High achiever		Set maths		Not set maths	
	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Teaching same-attainment groups												
No/some lessons (reference)												
Half of lessons	0.003	0.112	0.025	0.100	-0.028	0.085	0.004	0.154	0.090	0.104	-0.118	0.118
Almost every lesson	-0.033	0.088	0.080	0.115	-0.070	0.083	0.244	0.179	0.074	0.138	-0.075	0.117
Whole-class teaching												
No/some lessons (reference)												
Half of lessons	-0.065	0.113	-0.049	0.084	-0.055	0.089	-0.034	0.125	0.129	0.170	-0.119	0.110
Almost every lesson	0.022	0.139	0.037	0.090	0.043	0.102	0.176	0.117	0.186	0.165	0.010	0.131
Observations	**879**		**1,977**		**1,296**		**728**		**1,380**		**1,209**	

Table 25: The association between same-attainment grouping, mixed-attainment grouping, and mathematics achievement—Year 5 pupils, subgroup estimates

	FSM		Not FSM		Low achiever		High achiever		Set for maths		Not set for maths	
	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Teaching same-attainment groups												
No/some lessons (reference)												
Half of lessons	0.075	0.121	-0.051	0.100	-0.001	0.101	-0.073	0.164	-0.012	0.167	-0.105	0.109
Almost every lesson	-0.015	0.098	0.072	0.123	-0.050	0.088	0.300	0.175	0.039	0.138	-0.043	0.093
Teaching mixed-attainment groups												
No/some lessons (reference)												
Half of lessons	-0.111	0.178	0.092	0.088	-0.044	0.128	0.154	0.124	0.049	0.192	0.025	0.091
Almost every lesson	-0.006	0.126	-0.063	0.125	0.016	0.088	0.013	0.147	-0.138	0.155	0.100	0.097
Observations	**879**		**1,977**		**1,296**		**728**		**1,380**		**1,209**	

Robustness tests of the headline results for the primary outcomes are presented in Appendix D where either additional controls are added to the analysis model or multiple imputation is used to account for missing covariate data. Results from these robustness tests are consistent with those from the headline analysis model, with substantive conclusions largely unchanged. In particular, there continues to be little evidence of a clear association between the frequency that each teaching approach is used and the mathematics achievement of Year 5 pupils.

Trends in Mathematics and Science Study Year 9

The left-hand side of Table **26** investigates the association between mathematics achievement of Year 9 pupils and the frequency that they receive whole-class teaching and are taught in (within-class) attainment groups. Those pupils who are taught in attainment groups in almost every lesson do not achieve higher mathematics scores than their peers who are taught in same-attainment groups less frequently. The difference between the 'almost every' lesson and the 'no or some' lesson categorises is essentially zero (0.005 standard deviations). The estimated coefficients for the whole-class teaching variables are also small (effect sizes around 0.1 or less) and not statistically significant at conventional thresholds. Table **26** hence provides no evidence that regularly teaching Year 9 pupils in (within-class) attainment groups is associated with higher mathematics test scores.

Similarly, for the secondary outcome (pupils' self-confidence in mathematics), estimates are consistently small for both the frequency of teaching in same-attainment groups and whole-class teaching. There is hence no evidence that the frequency of either approach is associated with a substantial increase (or decrease) in Year 9 pupils' self-confidence in mathematics.

Table 26: The association between same-attainment grouping, whole-class teaching, and mathematics achievement amongst Year 9 pupils

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same-attainment groups						
No/some lessons (reference)						
Half of lessons	0.074	0.056	3288	-0.036	0.031	3288
Almost every lesson	0.005	0.055	3288	0.055	0.045	3288
Whole-class teaching						
No/some lessons (reference)						
Half of lessons	0.033	0.061	3288	0.060	0.032	3288
Almost every lesson	0.103	0.061	3288	0.010	0.026	3288

A similar conclusion is reached when one compares same-attainment and mixed-attainment groups for Year 9 pupils' self-confidence and achievement in mathematics (see Table **27**). Teaching in same-attainment groups in almost every lesson displays very little association with mathematics achievement (or mathematics self-confidence) compared to using this approach less frequently. Likewise, the estimated effect sizes are small (around 0.1 standard deviations or less) for the frequency of teaching in mixed-attainment groups and never approach statistical significance at conventional thresholds.

Table 27: The association between same-attainment grouping, mixed-attainment grouping, and mathematics achievement amongst Year 9 pupils

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same-attainment groups						
No/some lessons (reference)						
Half of lessons	0.017	0.087	3288	-0.041	0.060	3288
Almost every lesson	0.018	0.056	3288	0.049	0.044	3288
Teaching mixed-attainment groups						
No/some lessons (reference)						
Half of lessons	0.090	0.089	3288	0.004	0.052	3288
Almost every lesson	0.027	0.082	3288	-0.046	0.027	3288

Table 28 and Table 29 (page 37) illustrate whether these results differ across various subgroups. On the whole, there is little evidence that any of the three approaches are strongly associated with higher levels of mathematics achievement for any of the subgroups considered. Effect size estimates are generally small and do not approach statistical significance at conventional levels.

The only possible exception is with respect to whole-class teaching for FSM and low-achieving pupils (see Table 28). For Year 9 FSM pupils, whole-class teaching is associated with around a 0.2 standard deviation increase in TIMSS mathematics scores. After a Bonferroni adjustment has been made, this sits just outside the boundary of statistical significance at the 5% level (95% confidence interval ranging from -0.01 to 0.40). Similarly, for low-achieving pupils, the estimated effect size for the link between whole-class teaching and mathematics achievement (as measured by TIMSS scores) is 0.17, though with the Bonferroni-corrected confidence interval ranging from just below zero (-0.004) up to 0.34. This finding should hence be treated as indicative, with appropriate caution exercised.

Robustness tests of the headline results for the primary outcomes are presented in Appendix E where either additional controls are added to the analysis model or multiple imputation is used to account for missing covariate data. Results from these robustness tests are consistent with those from the headline analysis model, with substantive conclusions largely unchanged. In particular, there continues to be little overall association between within-class mixed-attainment grouping, same-attainment grouping or whole-class teaching and children’s mathematics achievement. Similarly, there remains little evidence of a link between the frequency that each teaching approach is used and the self-confidence of Year 9 pupils.

Table 28: The association between same-attainment grouping, whole-class teaching, and mathematics achievement—Year 9 pupils, subgroup estimates

	FSM		Not FSM		Low achiever		High achiever	
	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Teaching same-attainment groups								
No/some lessons (reference)								
Half of lessons	0.089	0.081	0.064	0.055	0.031	0.060	0.157	0.117
Almost every lesson	0.038	0.070	-0.004	0.056	0.017	0.093	0.005	0.079
Whole-class teaching								
No/some lessons (reference)								
Half of lessons	0.113	0.079	0.016	0.062	0.085	0.066	0.064	0.103
Almost every lesson	0.198	0.086	0.084	0.062	0.169	0.073	0.023	0.082
Observations	**787**		**2526**		**1111**		**1045**	

Table 29: The association between same-attainment grouping, mixed-attainment grouping, and mathematics achievement—Year 9 pupils, subgroup estimates

	FSM		Not FSM		Low achiever		High achiever	
	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Teaching same-attainment groups								
No/some lessons (reference)								
Half of lessons	0.077	0.127	-0.004	0.076	-0.045	0.075	0.014	0.218
Almost every lesson	0.034	0.077	0.013	0.057	0.059	0.100	0.004	0.068
Teaching mixed-attainment groups								
No/some lessons (reference)								
Half of lessons	0.017	0.132	0.105	0.077	0.153	0.094	0.182	0.194
Almost every lesson	-0.005	0.124	0.036	0.087	0.068	0.084	0.105	0.097
Observations	**787**		**2526**		**1111**		**1045**	

Conclusion

Key conclusions
There is no evidence of an association between within-class attainment grouping and Year 2 children's mathematics test scores.
There is no evidence that the frequency of using any of the teaching approaches (same-attainment groups, mixed-attainment groups, or using a whole-class approach) is associated with higher mathematics achievement amongst Year 5 or Year 9 pupils (that is, none of the approaches were found to be superior to the others).
There is little evidence that the frequency of using different teaching approaches (whole-class teaching, within-class same-attainment grouping, or within-class mixed-attainment grouping) is associated with Year 5 and Year 9 pupils' self-confidence in their mathematics abilities.

Interpretation

The EEF's Teaching and Learning Toolkit notes how within-class attainment grouping is thought to be positively associated with pupil achievement. However, it notes that the academic evidence on this matter is relatively weak, particularly for secondary pupils. Likewise, most studies have compared within-class attainment grouping to whole-class teaching, and not to other alternative approaches such as within-class grouping of pupils of mixed abilities.

Moreover, little is currently known about the practises and approaches actually in use in England. How frequently are pupils exposed to whole-class teaching as opposed to mixed-attainment or same-attainment grouping? Do teachers just use one single approach, or do they use a mix (for example, whole-class teaching for some or part of their lessons and getting pupils to work in attainment groups in others)? And are certain types of pupils exposed to (or do particular types of teachers use) one of these approaches more than the others?

This study has presented new evidence on this matter for England.

Year 2 pupils were found to be just as likely to be taught in (within-class) attainment groups regardless of their gender, ethnicity, family background, and prior levels of academic achievement. There was also little clear association between the characteristics of Year 2 class teachers and the use of within-class attainment grouping. Hence, the 'selection mechanism' does not seem to be related to observable teacher and pupil characteristics. However, within-class attainment grouping was found to be more common for Year 2 pupils in schools where the setting or streaming of pupils was *not* used. Hence school-management decisions—and, in particular, whether to use setting or streaming—does seem to be an important part of the 'selection mechanism'.

For Year 5 pupils, within-class mixed-attainment grouping was found to occur less frequently than either same-attainment grouping or whole-class teaching. Pupils taught in mixed-attainment groups in their mathematics class were less likely to be taught in same-attainment groups—with teachers seeming to trade-off between the two. There was some evidence, however, that teachers of Year 5 pupils used a mix of approaches: both the grouping of pupils (whether in same or mixed-attainment groups) and whole-class teaching were reported to be used relatively frequently. Little clear association was found between pupil characteristics and the frequency with which their mathematics teacher used different approaches, however, teacher characteristics did seem to play a part: less experienced teachers, for example, were less likely to teach in mixed-attainment groups. Thus, although observable teacher characteristics may be related to the decision of which teaching approach to use (the 'selection mechanism'), observable pupil characteristics are not.

For Year 9 pupils, whole-class teaching was found to be the most common approach to mathematics, while within-class mixed-attainment grouping of pupils was comparatively rare. This is likely due to most secondary schools using between-class attainment grouping (setting) for mathematics, limiting the amount of within-class variation in attainment. The use of whole-class teaching and group work (whether of mixed or same-attainment) amongst secondary teachers varied, with some teachers sticking exclusively to one approach while others mixed the two. Little evidence was found of teaching approaches differing substantially by pupil characteristics—with the important exception of prior achievement. Specifically, Year 9 pupils regularly taught in same-attainment groups within their class have higher Key Stage 2 scores than pupils rarely taught in same-attainment groups. On the other hand, Year 9 pupils in mathematics classes where mixed-attainment grouping is frequently used have lower prior achievement scores than their peers where this approach to teaching is uncommon. More experienced teachers were also found to use within-class mixed-attainment grouping less frequently while less experienced teachers were slightly less likely to use whole-class approaches. Consistent with the results for Year 5 pupils, observable teacher characteristics seem, to some extent, to be associated with the ‘selection mechanism’ of the frequency with which different teaching approaches are used.

No evidence was found that the use of within-class attainment grouping within Year 2 mathematics lessons is associated with higher (or lower) levels of mathematics achievement (on average). The same holds true for Year 2 pupils’ enjoyment of number work and reading test scores. Similarly, for Year 5 and Year 9 pupils, the frequency with which each teaching approach was used was not correlated with pupils’ mathematics outcomes. Likewise, throughout this report, there has been no clear evidence of an association between the frequency of different teaching approaches and pupil’s self-confidence in mathematics.

On the whole, there was no clear evidence that any of the three approaches were particularly beneficial for disadvantaged or low-achieving pupils’ progress in mathematics. The only possible exception—though the evidence continues to remain weak—is for whole-class teaching. In particular, disadvantaged and low-achieving Year 9 pupils where whole-class teaching was used frequently did experience slightly faster growth in their mathematics achievement, although there is quite a large degree of uncertainty surrounding this result. It is hence suggested that this is a specific area where further investigation and analysis is needed; it does not constitute a justification for a change in policy or practise.

How should these findings be interpreted? The largely null results of this study might indicate that teachers know themselves and their pupils best and that they use the approach that is most comfortable (and effective) for them. As the descriptive statistics indicated, a range of different approaches seem to be used, with some teachers tending to use one approach more frequently than others (for example, whole-class teaching rather than mixed-attainment grouping) while others seem to use a combination of different approaches. This possibly highlights the false dichotomy between approaches that is sometimes drawn, with the view that it is best to either use one approach or the other. Yet, in reality, teaching is more complex, with some topics potentially better suited to group work (or whole-class teaching) than others. Based upon the evidence presented here, it seems unwise to be too prescriptive about the approach teachers should favour (or how they should be mixed and matched together). The evidence is not strong enough to suggest that any single approach is better than any of the others, with teachers likely to be best placed to decide what works best for them and the classes that they teach.

Limitations

The clearest limitation of this study is that it is not based upon an experimental (or quasi-experimental) design. As noted throughout the report, estimates therefore refer to conditional associations only and are unlikely to capture cause and effect. In addition, although most of the effect sizes reported were not much larger (or smaller) than zero, they were often imprecisely estimated with quite large standard errors and wide confidence intervals. This reflects the limited sample size of the study, particularly in TIMSS, where the analysis is based upon the approaches used by a relatively small sample of teachers. It is also likely that the key variables used in the analysis—capturing the frequency of different approaches used by teachers—are measured with a degree of error. One cannot rule out the possibility that this is leading to attenuation of the estimates and is hence one potential explanation for the largely null results. It is also important to remember that the estimates presented only capture the practice of a single teacher during one particular year of children’s time at school. Finally, the data available within the MCS is from more than a decade ago and may not fully capture current practises within schools.

Future research and publications

The evidence presented in this report would be strengthened via replication of results using the upcoming TIMSS 2019 data. This would particularly help strengthen the evidence that frequent within-class attainment grouping of Year 9 pupils is positively associated with their mathematics achievement, as found in this report (based upon data from TIMSS 2015).

This report will form the basis for an academic paper, which will be submitted to an academic journal.

References

Archer, L., Francis, B., Miller, S., Taylor, B., Tereshchenko, A., Mazenod, A., Pepper, D. and Travers, M. (2018) 'The Symbolic Violence of Setting: A Bourdieusian Analysis of Mixed Methods Data on Secondary Students' Views About Setting', *British Educational Research Journal*, 44, pp. 119–140. DOI: 10.1002/berj.3321

Burke, M. and Sass, T. 'Classroom Peer Effects and Student Achievement', *Journal of Labor Economics*, 31 (1), pp. 51–82.

Coe, R., Jones, K., Searle, J., Kokotsaki, D., Kosnin, A. and Skinner, P. (2008) 'Evidence on the Effects of Selective Educational Systems': https://www.suttontrust.com/wp-content/uploads/2008/10/SuttonTrustFullReportFinal11.pdf

Connolly, P., Taylor, B., Francis, B., Archer, L., Hodgen, J., Mazenod, A. and Tereshchenko, A. (2019), 'The Misallocation of Students to Academic Sets in Maths: A Study of Secondary Schools in England', *British Educational Research Journal*, 45, pp. 873–897. DOI: 10.1002/berj.3530

Connolly, R. (2013) 'Interpreting Test Scores', Millennium Cohort Study Data Note 2013/1: https://cls.ucl.ac.uk/wp-content/uploads/2017/06/Data-Note-20131_MCS-Test-Scores_Roxanne-Connelly-revised.pdf

Education Datalab (2019) 'England's Schools Segregate by Attainment More Than Almost Every Other Country in the World': https://ffteducationdatalab.org.uk/2019/09/englands-schools-segregate-by-attainment-more-than-almost-every-other-country-in-the-world/

Francis, R., Archer, L., Hodgen, J., Pepper, D., Taylor, B. and Travers, M-C (2017) 'Exploring the Relative Lack of Impact of Research on "Attainment Grouping" in England: A Discourse Analytic Account', *Cambridge Journal of Education*, 47 (1), pp. 1–17. DOI: 10.1080/0305764X.2015.1093095

Francis, R., Taylor, R. and Tereshchenko, A. (2019) *Reassessing Attainment Grouping: Improving Practice for Equity and Attainment*. Routledge: Oxon.

Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D. and Shure, N. (2017) 'What Happens When Econometrics and Psychometrics Collide? An Example Using PISA Data', *Economics of Education Review*, 61, pp. 51–58.

Jiroutek, M. R. and Turner, J. R. (2018) 'Why It is Nonsensical to Use Retrospective Power Analyses to Conduct a Postmortem on Your Study', *Journal of Clinical Hypertension*, 20, pp. 408–410.

Ketende, S. and Jones, E. (2011) 'User Guide to Analysing MCS Data Using Stata': https://cls.ucl.ac.uk/wp-content/uploads/2017/07/User-Guide-to-Analysing-MCS-Data-using-Stata.pdf

Macdonald, K. (2008) 'PV: Stata Module to Perform Estimation with Plausible Values', Statistical Software Components S456951, Boston College Department of Economics, revised 03 Feb 2019.

Martin, M. O., Mullis, I. V. S. and Hooper, M. (eds) (2016) 'Methods and Procedures in TIMSS 2015', retrieved from Boston College, TIMSS and PIRLS International Study Center website: http://timssandpirls.bc.edu/publications/timss/2015-methods.html

Muijs, D. and Dunne, M. (2010) 'Setting by Attainment – or is it? A Quantitative Study of Determinants of Set Placement in English Secondary Schools', *Educational Research*, 52 (4), pp. 391–407. DOI: 10.1080/00131881.2010.524750

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.

Smith, C. and Sutherland, M. (2003) Setting or Mixed Attainment? Teachers' Views of the Organisation of Pupils for Learning', *Journal of Research in Special Educational Needs*, 3 (3), pp. 141–146.

Taylor, R., Francis, B., Archer, L., Hodgen, J., Pepper, D., Tereshchenko, A. and Travers, M-C. (2017) 'Factors Deterring Schools from Mixed Attainment Teaching Practice', *Pedagogy, Culture and Society*, 25 (3), pp. 327–345. DOI: 10.1080/14681366.2016.1256908

Appendix A. Characteristics of pupils included in and excluded from the analytic sample.

Appendix Table A1. Characteristics of pupils included and excluded from the MCS analytic sample

	In analytic sample		Not in analytic sample	
	n	% / mean	n	% / mean
Socio-economic background				
Family income	7,892	362	11,284	293
% father hold NVQ level 4 / 5	7,640	30%	10,912	20%
% mother hold NVQ level 4 / 5	7,640	35%	10,912	25%
Prior cognitive attainment				
Total FSP score (England only)	4,370	0.13	4,195	-0.14
Bracken school readiness score	6,664	0.06	7,179	-0.16
Age 5 pattern construction score	7,567	0.09	7,337	-0.09
Age 5 picture similarities	7,573	0.06	7,378	-0.06
Age 5 vocabulary score	7,583	0.12	7,378	-0.12
Demographics				
% White	7,618	88%	10,886	79%
% Male	7,640	50%	10,912	52%

TIMSS Year 5

Appendix Table A2. Characteristics of pupils included and excluded from the Year 5 analytic sample

	Year 5	
	Included in analytic sample	Excluded from analytic sample
Child characteristics		
% male	48%	51%
% FSM	31%	34%
% immigrant	9%	22%
% English not spoken at home	16%	32%
% regularly absent from school	90%	90%
Prior achievement		
Key Stage 1 maths score	16.3	16.2
Key Stage 1 reading score	15.7	15.6
Teacher characteristics		
% Subject specialism in maths	11%	18%
Years of teaching experience	10.3	9.7
Class characteristics		
Class size	27.7	25.7
% students with language issues	8%	14%
Minutes per week teaching class	303	283
Average KS1 maths scores of class	16.3	16.1
Standard deviation of class KS1 maths scores	3.0	2.8
School characteristics		
Pupils set for mathematics	57%	47%
0-10% of pupils from poor backgrounds	19%	16%
11-25% of pupils from poor backgrounds	33%	24%
26-50% of pupils from poor backgrounds	20%	12%
51-100% of pupils from poor backgrounds	19%	32%
Resource shortages affecting maths instruction scale	-0.04	0.21
Observations	**2,856**	**1,150**

TIMSS Year 9

Appendix Table A3. Characteristics of pupils included and excluded from the Year 9 analytic sample

	Year 9	
	Included in analytic sample	Excluded from analytic sample
Child characteristics		
% male	49%	54%
% FSM	26%	28%
% immigrant	7%	15%
% English not spoken at home	4%	9%
% regularly attend school	92%	93%
Prior achievement		
Key Stage 2 maths score	70.6	70.0
Key Stage 2 reading score	31.7	31.4
Teacher characteristics		
Years of teaching experience	10.3	8.8
Class size	26.5	25.9
Class characteristics		
% students with language issues	3%	3%
Minutes per week teaching class	198.7	194.1
Average KS2 maths scores of class	70.5	68.2
Standard deviation of class KS2 maths scores	10.0	9.7
School characteristics		
Pupils set for mathematics	99%	100%
0-10% of pupils from poor backgrounds	17%	13%
11-25% of pupils from poor backgrounds	36%	32%
26-50% of pupils from poor backgrounds	21%	13%
51-100% of pupils from poor backgrounds	10%	9%
Missing data on pupils from poor background	16%	32%
Resource shortages affecting maths instruction scale	-0.02	0.06
Observations	**3288**	**1526**

Appendix B. Associations between attainment grouping and outcomes measures.

Appendix Table B1. The association between within-class attainment grouping, enjoyment of number work and reading achievement amongst Year 2 pupils. Sub-group estimates

	Enjoy number work			Reading scores		
	Effect	SE	n	Effect	SE	n
Family-income						
Low-income (bottom third)	-0.145	0.140	2,450	-0.018	0.053	2,570
High-income (top third)	0.043	0.117	2,517	-0.018	0.053	2,578
Prior achievement						
Low prior achievement (bottom third)	0.031	0.129	2,354	-0.128	0.052	2,475
High prior achievement (top third)	-0.043	0.132	2,432	0.049	0.065	2,497
Setting/streaming used for subject						
No	-0.074	0.115	4,249	-0.045	0.045	4,403
Yes	0.105	0.107	2,684	-0.034	0.044	2,768

Appendix Table B2. The association between same-attainment grouping, whole-class teaching and mathematics self-confidence amongst Year 5 pupils. Sub-group estimates.

	FSM		Not FSM		Low achiever		High achiever		Set maths		Not set maths	
	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Teaching same attainment groups												
No/some lessons (reference)												
Half of lessons	-0.042	0.113	0.010	0.045	0.107	0.099	-0.107	0.059	-0.017	0.079	0.098	0.075
Almost every lesson	-0.030	0.081	0.020	0.048	0.120	0.060	-0.085	0.051	0.009	0.063	0.122	0.074
Whole class teaching												
No/some lessons (reference)												
Half of lessons	0.229	0.142	0.021	0.036	0.127	0.078	-0.065	0.038	0.087	0.076	0.070	0.042
Almost every lesson	0.129	0.088	0.054	0.044	0.119	0.070	-0.034	0.032	0.089	0.056	0.068	0.061
Observations	**879**		**1,977**		**1,296**		**728**		**1,380**		**1,209**	

Appendix Table B3. The association between same-attainment grouping, mixed-attainment grouping and mathematics self-confidence amongst Year 5 pupils. Sub-group estimates.

	FSM		Not FSM		Low achiever		High achiever		Set for maths		Not set for maths	
	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Teaching same attainment groups												
No/some lessons (reference)												
Half of lessons	0.009	0.152	0.044	0.048	0.140	0.120	-0.089	0.054	-0.049	0.076	0.141	0.092
Almost every lesson	0.025	0.091	0.017	0.050	0.136	0.071	-0.094	0.056	0.006	0.064	0.133	0.084
Teaching mixed attainment groups												
No/some lessons (reference)												
Half of lessons	-0.015	0.107	-0.080	0.054	-0.057	0.066	-0.046	0.041	0.031	0.064	-0.077	0.049
Almost every lesson	0.196	0.172	-0.056	0.035	0.014	0.061	-0.068	0.048	0.012	0.066	0.032	0.048
Observations	**879**		**1977**		**1296**		**728**		**1380**		**1209**	

Appendix Table B4. The association between same-attainment grouping, whole-class teaching and mathematics self-confidence amongst Year 9 pupils. Sub-group estimates.

	FSM		Not FSM		Low achiever		High achiever	
	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Teaching same attainment groups								
No/some lessons (reference)								
Half of lessons	-0.054	0.045	-0.042	0.043	-0.006	0.060	-0.032	0.039
Almost every lesson	0.041	0.048	0.052	0.056	0.237	0.149	-0.017	0.028
Whole class teaching								
No/some lessons (reference)								
Half of lessons	0.096	0.073	0.046	0.037	0.061	0.066	-0.019	0.025
Almost every lesson	0.017	0.036	-0.001	0.038	0.075	0.059	-0.047	0.046
Observations	**787**		**2,526**		**1,111**		**1,045**	

Appendix Table B5. The association between same-attainment grouping, mixed-attainment grouping and mathematics self-confidence amongst Year 9 pupils. Sub-group estimates.

	FSM		Not FSM		Low achiever		High achiever	
	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Teaching same attainment groups								
No/some lessons (reference)								
Half of lessons	-0.126	0.103	-0.022	0.070	0.048	0.097	-0.250	0.202
Almost every lesson	0.051	0.053	0.041	0.055	0.226	0.151	-0.010	0.031
Teaching mixed attainment groups								
No/some lessons (reference)								
Half of lessons	0.129	0.117	-0.037	0.049	-0.091	0.084	0.272	0.218
Almost every lesson	-0.034	0.044	-0.070	0.048	-0.049	0.036	-0.003	0.033
Observations	**787**		**2,526**		**1,111**		**1,045**	

Appendix C. Robustness tests for Year 2 MCS results.

<u>Alternative estimates restricting analytic sample to those who completed the questionnaire before September 2008</u>

Appendix Table C1. The association between within-class attainment grouping and mathematics achievement amongst Year 2 pupils (primary outcome).

	Mathematics scores		
	Effect	SE	N
Within-class attainment grouping used?			
No (Reference group)			
Yes	0.049	0.046	3296

Appendix Table C2. The association between within-class attainment grouping, enjoyment of number work and reading achievement amongst Year 2 pupils (secondary outcomes).

	Enjoy number work			Reading scores		
	Logit	SE	n	Effect	SE	n
Within-class attainment grouping used?						
No (Reference group)						
Yes	0.113	0.122	3137	-0.046	0.043	3,239

<u>Alternative estimates based upon propensity score matching</u>

Appendix Table C3. Matching estimates of the association between within-class attainment grouping and mathematics achievement amongst Year 2 pupils (primary outcome).

	Mathematics scores		
	Effect	SE	N
Within-class attainment grouping used?			
No (Reference group)			
Yes	0.063	0.068	7,898

Appendix Table C4. Matching estimates of the association between within-class attainment grouping, enjoyment of number work and reading achievement amongst Year 2 pupils (secondary outcomes)

	Enjoy number work			Reading scores		
	Probattainment difference	SE	n	Effect	SE	n
Within-class attainment grouping used?						
No (Reference group)						
Yes	-4.6%	1.8%	7,488	0.017	0.049	7,741

Appendix D. Robustness tests for Year 5 TIMSS results.

Alternative estimates including additional controls

Appendix Table D1. The association between same-attainment grouping, whole-class teaching and mathematics achievement amongst Year 5 pupils. Robustness test including additional controls.

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same attainment groups						
No/some lessons (reference)						
Half of lessons	0.011	0.078	2851	0.008	0.043	2851
Almost every lesson	0.019	0.085	2851	0.023	0.039	2851
Whole class teaching						
No/some lessons (reference)						
Half of lessons	-0.054	0.067	2851	0.076	0.042	2851
Almost every lesson	0.023	0.083	2851	0.067	0.042	2851

Appendix Table D2. The association between same-attainment grouping, mixed-attainment grouping and mathematics achievement amongst Year 5 pupils. Robustness test including additional controls.

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same attainment groups						
No/some lessons (reference)						
Half of lessons	-0.059	0.078	2851	0.042	0.051	2851
Almost every lesson	0.002	0.098	2851	0.025	0.043	2851
Teaching mixed attainment groups						
No/some lessons (reference)						
Half of lessons	0.086	0.078	2851	-0.066	0.040	2851
Almost every lesson	-0.083	0.118	2851	-0.010	0.055	2851

Appendix Table D3. The association between same-attainment grouping, whole-class teaching and mathematics achievement amongst Year 5 pupils. Robustness test using multiple imputation.

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same attainment groups						
No/some lessons (reference)						
Half of lessons	0.046	0.068	3587	-0.015	0.046	3587
Almost every lesson	0.045	0.069	3587	0.019	0.050	3587
Whole class teaching						
No/some lessons (reference)						
Half of lessons	-0.011	0.067	3587	0.062	0.036	3587
Almost every lesson	0.075	0.070	3587	0.079	0.040	3587

Appendix Table D4. The association between same-attainment grouping, mixed-attainment grouping and mathematics achievement amongst Year 5 pupils. Robustness test using multiple imputation.

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same attainment groups						
No/some lessons (reference)						
Half of lessons	0.047	0.070	3587	0.019	0.050	3587
Almost every lesson	0.068	0.072	3587	0.035	0.052	3587
Teaching mixed attainment groups						
No/some lessons (reference)						
Half of lessons	0.008	0.071	3587	-0.051	0.043	3587
Almost every lesson	0.012	0.076	3587	0.018	0.055	3587

Appendix E. Robustness tests for Year 9 TIMSS results.

Alternative estimates including additional controls

Appendix Table E1. The association between same-attainment grouping, whole-class teaching and mathematics achievement amongst Year 9 pupils. Robustness test including additional controls.

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same attainment groups						
No/some lessons (reference)						
Half of lessons	0.044	0.057	3285	-0.028	0.031	3285
Almost every lesson	0.015	0.054	3285	0.051	0.045	3285
Whole class teaching						
No/some lessons (reference)						
Half of lessons	0.028	0.058	3285	0.067	0.032	3285
Almost every lesson	0.080	0.057	3285	0.019	0.026	3285

Appendix Table E2. The association between same-attainment grouping, mixed-attainment grouping and mathematics achievement amongst Year 9 pupils. Robustness test including additional controls.

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same attainment groups						
No/some lessons (reference)						
Half of lessons	-0.015	0.085	3285	-0.029	0.060	3285
Almost every lesson	0.028	0.056	3285	0.047	0.045	3285
Teaching mixed attainment groups						
No/some lessons (reference)						
Half of lessons	0.089	0.086	3285	0.000	0.052	3285
Almost every lesson	0.007	0.081	3285	-0.041	0.028	3285

Appendix Table E3. The association between same-attainment grouping, whole-class teaching and mathematics achievement amongst Year 9 pupils. Robustness test using multiple imputation.

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same attainment groups						
No/some lessons (reference)						
Half of lessons	0.064	0.051	4348	-0.056	0.037	4348
Almost every lesson	-0.009	0.048	4348	0.026	0.039	4348
Whole class teaching						
No/some lessons (reference)						
Half of lessons	0.030	0.054	4348	0.039	0.039	4348
Almost every lesson	0.100	0.052	4348	0.010	0.032	4348

Appendix Table E4. The association between same-attainment grouping, mixed-attainment grouping and mathematics achievement amongst Year 9 pupils. Robustness test using multiple imputation.

	Maths achievement			Self-confidence		
	Effect	SE	n	Effect	SE	n
Teaching same attainment groups						
No/some lessons (reference)						
Half of lessons	0.012	0.059	4348	-0.041	0.051	4348
Almost every lesson	0.006	0.051	4348	0.018	0.041	4348
Teaching mixed attainment groups						
No/some lessons (reference)						
Half of lessons	0.094	0.058	4348	-0.030	0.044	4348
Almost every lesson	0.071	0.062	4348	-0.073	0.029	4348

Appendix F: Questions used to gather information on within-class attainment grouping in the MCS

Class groupings

Some schools group children within the same class by general attainment and they are taught in these attainment groups for most or all lessons. We refer to this as within-class attainment grouping.

Some schools group children within the same class by attainment for certain subjects only and they may be taught in different attainment groups for different subjects. We refer to this as within-class subject grouping.

Other schools do not group children by attainment within classes. Some schools may use within-class groupings in addition to between class streaming and setting and others may use within-class groupings instead of between class streaming and setting.

Some schools may not use any general or subject specific attainment groupings either within or between classes.

- Q55. In this child’s class, is there within-class attainment grouping? (Yes / No)
- Q56. How many within-class attainment groups are there? (Open text)
- Q57. Which group is this child in? (Highest, middle, lowest)
- Q58. In this child’s class, are there within-class subject groups for literacy? (Yes/No)
- Q59. How many within-class subject groups are there for literacy? (Open text)
- Q60. Which group is this child in for literacy? (Highest, middle, lowest)
- Q61. In this child’s class, are there within-class subject groups for maths? (Yes/No)
- Q62. How many within-class subject groups are there for maths? (Open text)
- Q63. Which group is this child in for maths? (Highest, middle, lowest)

This document is available for download at **https://educationendowmentfoundation.org.uk**

Education Endowment Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn