


**Please cite the Published Version**

Klss, Zsolt, Morris, Stephen  and Stefanelli, Alberto (2022) Assessment of changes to the membership of mathematics classes around Year 9, 10 and 11. Research Report. Education Endowment Foundation.

**Publisher:** Education Endowment Foundation

**Version:** Published Version

**Downloaded from:** <https://e-space.mmu.ac.uk/629256/>

**Usage rights:**  Open Government Licence 3.0

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)



Education  
Endowment  
Foundation

# Assessment of changes to the membership of mathematics classes around Year 9, 10 and 11

EEF Research Paper  
February 2022

Dr Zsolt Kiss, ZK Analytics  
Prof Stephen Morris, Manchester Metropolitan University  
[Alberto Stefanelli](#), ZK Analytics and KU Leuven

Simon Woodhead, Eedi  
Iris Hulls, Eedi



**I= Eedi**



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](mailto:jonathan.kay@eefoundation.org.uk)



[www.educationendowmentfoundation.org.uk](http://www.educationendowmentfoundation.org.uk)



# Contents

<b>Executive summary .....</b>	<b>3</b>
<b>1. Introduction.....</b>	<b>4</b>
<b>2. Within-year changes to class membership.....</b>	<b>7</b>
<b>3. Between-year changes to class membership .....</b>	<b>12</b>
<b>4. Recommendations.....</b>	<b>17</b>
<b>Appendix A: Focused literature review summary .....</b>	<b>19</b>
<b>Appendix B: Technical Note.....</b>	<b>22</b>
<b>References .....</b>	<b>32</b>

## Executive summary

The purpose of this study was to address the deficit in our knowledge around the prevalence of changes in class membership and its potential consequences for the design of trials and the analysis of trial data.

Much educational data is hierarchical with pupils nested in classes, classes in schools, and schools often in higher-level hierarchies. Hierarchical models are often used to account for the clustering at the various levels. In cases where class is included and controlled for in analysis it is usually done on the basis of class membership at a single point in time, at randomisation. However, if pupils move between classes, the points in time when the class membership is measured might become significant.

This research project draws on data derived during the course of the EEF-funded trial Diagnostic Questions<sup>1</sup> and includes records for a cohort of pupils as they moved through Years 9, 10, and 11 between 2017 and 2020. Overall, 25,668 unique pupils from 144 schools are included in the analysis with the key objectives of:

- understanding the extent of class to class moves within each academic year over the course of instruction in mathematics in Years 9, 10, and 11, as well as between Years 9 and 10 and Years 10 and 11; and
- understanding whether the move patterns identified are associated with any school or pupil characteristics.

The results show that:

- changes to class membership (that is, pupil moves between classes) do occur within academic years but their prevalence is relatively low;
- changes to class membership between academic years are a lot more prevalent as well as more complex as both class name and class membership need to be considered; we identified four scenarios:
  - similar class composition and the same name;
  - similar class composition and a different name;
  - different class composition and the same name;
  - different class composition and a different name; and
- there is little evidence to support a strong case for an association between moves and school or pupil characteristics.

We conclude that while the potential for moves should be always considered, in the within-year scenario the low prevalence of meaningful changes to class membership suggests that the logistical effort required to collect data on class membership several times across the year might not be warranted. Nonetheless, practitioners should take note of the difficulties associated with correctly identifying classes (due primarily to inconsistencies in class naming, duplication, and errors).

If we consider studies spanning two or more school years, between-year changes to class membership should be taken into account. Given the insights uncovered in this research, we recommend that, if logistically feasible, data records capturing class membership for sampled pupils are updated for the position at the beginning of each school year over the study period and further suggest that sensitivity analysis is performed to assess how far inferences are robust to the precise point in time class membership is taken into account in analysis. On a practical level, we highlight the difficulty of identifying class moves given that both class composition and class name need to be taken into consideration.

---

<sup>1</sup> <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/diagnostic-questions>

# 1. Introduction

## 1.1 Background

Much educational data is hierarchical. Pupils can be nested in classes, classes in schools, and schools often in higher-level hierarchies. Pupil outcomes are known to be dependent within these levels (EEF, 2018). To control for this, substantial analytical efforts are made in the analysis of educational data. Such efforts are evident in the analysis of data from school-based trials where hierarchical and other models are used to account for the clustering at the various levels.

Such analyses, however, often ignore the class level. In May 2019, the EEF published a research paper that explores whether the classroom level matters in the design of educational trials. Demack (2019, p.3) argues that ‘setting/streaming creates within-school attainment clusters of pupils that acts in conjunction with other factors (such as “the teacher”) to make pupils within one class “similar” to each other and less similar to pupils in other classes. This class-level clustering is a structural reality of the data in many educational trials regardless of whether it is acknowledged in the research design.’ The study concludes that omitting the class level in the design of experiments is likely to overstate their design sensitivity.

In rare cases where class is included and controlled for in analysis, it is usually done on the basis of class membership at a single point in time; that is, at randomisation. However, a pupil’s class membership might not be fixed over the course of the trial. Pupils can change classes within a school year as well as between school years where studies extend in length for periods of longer than a single school year. However, there is very little evidence on the extent to which class membership changes or when they are more likely to occur, that is, within or between school years.

Statistical methods such as hierarchical linear modelling or general estimating equations can be used in estimation of variances that account for clustering at the level of the class. However, if class membership is not fixed<sup>2</sup> and students are transferred between classes over the course of a study, the question becomes: at what time point should class membership be captured? Should this be done at the point of randomisation, at some intermediate point, or the point at which end-point outcome measurements are taken?

The primary purpose of this study is to address the deficit in our knowledge around the prevalence of changes in class membership and its potential consequences for the design of trials and the analysis of trial data.

## 1.2 Project objectives

This research project aims to extend the evidence base pertaining to how pupils move between maths classes in Years 9, 10, and 11. We focus our efforts on understanding:

- the extent of class to class moves among pupils, relative to their starting position, within each academic year over the course of instruction in GCSE maths (Years 10 and 11) as well as Year 9;
- the extent of class moves between Years 9 and 10 and between Years 10 and 11;
- whether patterns in class moves vary by type of school, free school meal status (FSM), gender, and month of birth;
- the extent to which grouping of pupils (such as setting or streaming) is related to class moves; and
- the potential consequences for the design and analysis of trials of including the class level captured at different points in time.

---

<sup>2</sup> Schools use streaming or setting to determine class membership and as pupils’ performance changes, so might their class membership. Moreover, such moves are likely to be correlated with attainment.

### 1.3 Data sources

The analysis draws on a unique dataset derived during the course of the EEF-funded trial Diagnostic Questions.<sup>3</sup> The analysis draws on data recorded by Eedi through the implementation of the Diagnostic Question platform in participating schools.

Diagnostic Questions is an online assessment tool offered by Eedi to schools and teachers. It enables teachers to assign diagnostic quizzes to students. Each question in the quiz is multiple choice with exactly four answers, one correct and three distractors that correspond to unique misconceptions. The system gives instant feedback and teachers can use feedback to direct the focus of their teaching. In the trial, the platform is used by maths teachers in schools, with tasks set as homework.

As part of providing this service, Eedi collects and stores administrative data on students and schools. Eedi is integrated with the Wonde<sup>4</sup> system that offers information drawn from the schools' Management Information System (MiS) of the participant schools. This data contains information on the make-up of classes over time and time-stamps any changes in class composition with the information being available for all subjects in a school.

The dataset includes records for a cohort of pupils as they move through Years 9, 10, and 11 between 2017 and 2020. After restricting the data capturing pupil membership of maths classes, we identified 81,480<sup>5</sup> records referring to 25,668<sup>6</sup> unique pupils across 144 schools<sup>7</sup>. There are:

- 20,168 unique pupils with records in both Years 9 and 10; and
- 20,652 unique pupils with records in both Years 10 and 11.

The records are structured by class membership. Each row in the dataset represents one instance of a class membership for a pupil, specifying the date when the pupil was added to a class, the class size, the class code,<sup>8</sup> and the subject.<sup>9</sup> Several records per pupil per year suggest potential changes in class membership.

In addition to the Eedi dataset we use school-level data made available by the DfE (for example, School Census, attainment statistics, and so forth). This allows us to assess any association between moves and school characteristics.

### 1.4 Data analysis

The analysis was implemented in successive stages.

First, we carried out a focused literature review to understand the evidence (in EEF trials and beyond) pertaining to the statistical treatment and modelling of the class level in experimental as well as observational studies. If clustering by class was accounted for by the research, we were interested in understanding the decisions around the time point at which class membership was established. We aimed for this review (attached as Appendix 1) to frame the analytical strategy to be adopted in this study. The existing relevant literature was scant to non-existing. Furthermore, it appears

---

<sup>3</sup> <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/diagnostic-questions>

<sup>4</sup> <https://www.wonde.com/>

<sup>5</sup> This decreases to 79,991 once records with incorrect dates are removed.

<sup>6</sup> This decreases to 25,651 once records with incorrect dates are removed.

<sup>7</sup> Seventy-three schools were part of the treatment group in the EEF trial and 71 were in the control group.

<sup>8</sup> This is the code a class is registered with in the school's Management System and it is unique to each school (for example, '9S/Ma3').

<sup>9</sup> In instances where the subject field was empty the subject was determined by relying on the class code.

that nearly all studies omit class membership from their statistical analyses and we were able to identify several key reasons for this, highlighted in Appendix 1.

The analysis uses individual-level anonymised data collected by Eedi and only includes schools that were part of the EEF-funded Diagnostic Questions trial.

In the first phase of the research, we carried out data diagnostics and performed data cleaning to ensure a consistent analytical dataset. As part of this process we:

- developed indicator variables capturing whether a school used a form of grouping;<sup>10</sup>
- extracted class subject from class codes; and
- split the data into the component school years.

The second phase of the analysis focused on the development of an algorithm to objectively identify changes in class membership and estimate the incidence of such changes. Consistent with our research aims, we measured within- and between-year changes separately.

The third phase of the research assessed whether the changes to class membership were related to pupil characteristics such as FSM, gender, month of birth, or type of school as well as school characteristics (including the extent grouping was used within the school, past school-level attainment at KS4, teacher to pupil ratio, type of school, and school-level values for FSM, EAL, and SEN). Regression modelling was used to examine the associations between these variables and class moves. Please refer to the Technical Note section of this report for further details.

## 1.5 Representativeness

We assessed the extent to which the sample of schools was representative of secondary schools in England. Table 1 compares our sample with the population of schools on key characteristics. The comparison suggests that there are only minimal differences on key variables such as attainment, school size, and FSM. As such, the analysis is not weighted.

Table 1: Comparison of the sample with the population of secondary schools in England

Characteristic	Analytical sample*	Population (January 2018)
Attainment in English and maths (9-5)	43.1%	43.3% <sup>11</sup>
Average school size	1,008	948 <sup>9</sup>
Average class size	24.3	21.2 <sup>12</sup>
Pupils eligible for FSM	12.5%	12.4% <sup>9</sup>
EAL	13.8%	16.6% <sup>9</sup>
SEN	10.5%	14.6% <sup>13</sup>

\* Figures calculated as the weighted average across the school included in the analysis.

<sup>10</sup> This was a complex process and we discuss it in Appendix 2, Technical Note.

<sup>11</sup>[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/774014/2018\\_KS4\\_main\\_text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/774014/2018_KS4_main_text.pdf)

<sup>12</sup>[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/719226/Schools\\_Pupils\\_and\\_their\\_Characteristics\\_2018\\_Main\\_Text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/719226/Schools_Pupils_and_their_Characteristics_2018_Main_Text.pdf)

<sup>13</sup>[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/729208/SEN\\_2018\\_Text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/729208/SEN_2018_Text.pdf)

## 2. Within-year changes to class membership

### 2.1 Defining classes and moves

In theory, the concept of a ‘move’ is relatively straightforward; it can be thought as the transfer of a pupil from one class to another. However, in practice, developing a robust and consistent definition that can be operationalised in our dataset is considerably more complicated and hinges on how the ‘class’ is defined. For the purpose of assessing within-year changes in class membership, we define the class using a nominal definition:

*A class is defined by its name. Classes are recorded and identified in School Management Systems using codes that are unique to each class within a given school. Usually the code identifies the year, teacher, subject, and group. On a practical level the codes are the only means through which the group of pupils studying together can be identified.*

To put it simply, a move occurs within an academic year when a pupil has more than one class membership during that year. Using the nominal definition outlined above, in practice, a move occurs when a pupil transfers from a group identified by a certain class code to a group identified by a different code during the course of the academic year (for example, from 9S/Ma1 to 9S/Ma2). However, it is important to note that our definition is not absolute. There are situations where even if such a change occurs, we do not consider it a move: for example, when the second record is generated within seven days after the first record. We discuss this further in Section 2.5.2.2.

We can identify several instances of class membership for some pupils in each academic year, suggesting the possibility of within-year moves between classes. Table 2 illustrates the proportion of pupils who have multiple records with apparently different class memberships.

Table 2: Frequency of multiple class membership records

	Year 9	Year 10	Year 11
Percentage of pupils with more than one class membership	14%	16%	20%

At first glance, it would be convenient to assume that each instance where a student has several memberships represents a move. This would be an error. As we detail in the following sections, our analyses suggest that not all instances of multiple records represent moves between classes. In addition to moves, we find that the existence of multiple membership records can be explained by:

- duplication of class names;
- consecutive records occurring within seven days of each other; and
- complex instances that cannot be categorised within contextual knowledge.

## 2.2 Prevalence of class moves

The in-depth assessment of class membership records for pupils with multiple records (see Table 2 above) suggests that there are indeed pupils who change classes within an academic year. Table 3 summarizes the likely range of moves, once other explanations for multiple records are discounted.

Table 3: Range of class moves

Academic Year	Prevalence of within-year moves
Year 9	Below 1%
Year 10	4% to 10%
Year 11	5% to 17%

## 2.3 Pupil- and school-level characteristics associated with within-year changes

### 2.3.1 School characteristics

We carried out regression analysis to assess whether within-year moves are linked to school characteristics. We only implemented models for Years 10 and 11, given the very low incidence of moves in Year 9.

The analysis was carried out at the school level. The outcome was a continuous variable indicating the number of within-year moves in each school.<sup>14</sup>

The results are included in Appendix 2 and show that while the model has a non-trivial fit for Year 10, the fit of the model run on Year 11 data is particularly poor. As a result, we will not discuss the results produced for Year 11. The poor fit could potentially be explained by our inability to more accurately identify patterns of moves, given the more complex structure of class in Year 11. Alternatively, this could also be explained by the genuine randomness of the moves in Year 11.

The model for Year 10 suggests that the number of class moves were:

- negatively associated with the percentage of SEN pupils;
- positively associated with the percentage of FSM pupils; and
- negatively associated with the number of pupils grouped by attainment (although this grouping only reached statistical significance at the 10% level), however, this could be a function of the fact that the variable is derived based on the empirical analysis of class membership patterns across maths and English classes; Appendix B details how this measure was derived.

### 2.3.2 Pupil characteristics

Analysis was also carried out to determine whether pupil characteristics (gender, FSM, month of birth) were associated with class moves. The analysis was only carried out for the pupils in schools that were part of the treatment group in the initial trial. This was because we did not hold data on the pupils in the control schools.

The outcome used was a dichotomous variable indicating whether the pupil moved classes. The analysis was implemented in two forms: a logistic regression with fixed effects for schools and as a mixed-effects logistic regression

---

<sup>14</sup> An alternative analysis was also implemented whereby the outcome was recoded into a dichotomous variable indicating whether any moves were identified in a school. Logistic regression was implemented to assess whether schools that experienced moves were different compared those that did not. The analysis produced similar results to the ones described in the text.

with random intercepts for schools. The results of the models are consistent with one another and overall suggest that class moves are not correlated with (observable) pupil characteristics.

## 2.4 Relevance to EEF trials

In nearly 60% of EEF trials,<sup>15</sup> schools are randomised to intervention and control conditions but outcomes are measured at the pupil level, with effect sizes reported for pupils. These pupils are grouped within classes and, particularly in secondary school settings, pupils are often sorted into classes based on prior attainment and/or teacher judgement as to their level. Ideally, therefore, regression models from which estimated effects are obtained should take account of clustering at both the class and school level. In practice, models often account for clustering at the school level, only because of difficulties obtaining information about class membership within schools. Such information is often expensive to obtain, burdensome for schools, and of poor quality when it is available.

For EEF-funded studies where focal pupil cohorts are Year 10 or Year 11—5% of all trials according to a recent review<sup>15</sup>—and where access to information about class membership is available, the accepted practice has been to account for class membership at the commencement of the trial. Where studies extend over one school year, based on the results we obtain here, this would seem a reasonable strategy because it appears that there is only a small amount of class to class movement within a school year. Thus, the majority of pupils start and finish the school year in the same class. As a result, there appears to be little to gain from collecting further information on class membership as the school year unfolds and using this information in statistical modelling.

## 2.5 Data processing and determination of the prevalence of moves

### 2.5.1 Computing the minimum incidence

The minimum incidence was included in Table 3 as the lower end of the range. This figure is based on a conservative analysis of the data, whereby we defined moves on the basis of an automated and objective algorithm. The results are detailed in Table 4.

Table 4: Minimum incidence of pupils changing classes during the course of the academic year

Academic year	Pupils	Moves between classes	Moves between classes as proportion of pupils	Average moves per school (SD) <i>as a proportion of the entire sample of schools</i>	Number of schools with at least one move (total number of schools)	Average moves per school (SD) <i>as a proportion of the schools with moves</i>
Year 9	21,578	66	0.3%	0.5 (5)	4 (121)	17 (26)
Year 10	23,888	1,175	5%	8 (16)	78 (138)	13 (18)
Year 11	21,394	894	4%	6 (12)	65 (130)	12 (14)

Table 5 illustrates how moves can be detected. The example uses real but anonymised data: the identities of the pupils and schools have been replaced with digits (1, 2) and letters (A, B).

As discussed earlier, we define a move as an instance in which a pupil changes between two maths classes. In this context, we define a class using its name: this is the code assigned to the class by the school as can be seen in the 'class code' column. In the example below, we can see that both pupils were assigned to a class at the start of the

<sup>15</sup><https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-evaluation-reports-and-research-papers/syntheses-of-eef-evaluations/review-of-eef-projects>

academic year (maths class 4 for each student). Later in the year (January in the case of both pupils), each pupil was registered with a different class. Pupil 1 was ‘moved’ from class 4 to class 5, while pupil 2 was ‘moved’ from class 4 to class 7.

Table 5: Examples of moves between classes within Year 11

Pupil	School	Year	Class code	Date registered in class	Class size
1	A	11	11ta/Ma4	11/09/2019	35
1	A	11	11ta/Ma5	16/01/2020	31
2	B	11	MAT11X-4-ADP	17/09/2019	23
2	B	11	MAT11X-7-ADP	07/01/2020	21

## 2.5.2 Computing the maximum incidence

Due to differences and irregularities in the naming of classes as well as possible human errors in the entry of class codes in different schools, we are aware that some moves may not be identified by simply using the automated algorithm (see Table 6, ‘complex cases’ column). Once these instances are also taken into account as *possible* moves, the *maximum* potential incidence for moves increases to 1% for Year 9, 10% for Year 10, and 17% for Year 11. However, it is likely that only a some of these instances are actual moves.

In the process of identifying moves between classes, a vast data processing effort was needed. This involved identifying the various types of patterns that exist in the way class membership is recorded and stored in school management systems and how it can be accessed and extracted via Wonde. Table 6 summarises the types of patterns we identified and accounted for to obtain an unbiased minimal measure of moves as well as the complex cases that we use as the upper bounds.

In addition to the moves between classes, we also identify instances where:

- class names are duplicated;
- different class memberships are recorded in quick succession (within seven days of each other); and
- unknown—complex instances that cannot be categorised using rule-based allocation.

Table 6: Class membership patterns within years

Academic year	Pupils	Pupils registered in classes with duplicate class names	Pupils registered in a class within seven days of first record	Pupils who move between classes	Pupils whose class membership qualifies as a complex case
Year 9	21,578	2,874	397	66	206
Year 10	23,888	758	714	1,175	1,079
Year 11	21,394	1,460	1,038	894	2,762

### 2.5.2.1 Duplicate class names

We find that there are instances where several records pertaining to the same pupil reference the same class name. In some of these cases the only difference in the class name between the records is whether the first character is capitalised or not. The examples included in Table 5 show that pupil 3 was registered in two classes, one named ‘Ma1’ while the second was named ‘ma1’. Similarly, pupil 4 was registered in three classes, two of them named ‘Ma4’ and one ‘ma4’.

Based on these patterns, two distinct situations emerge:

- *Class names are entirely identical.* Approximately 70% of cases with duplicate class names in Years 9 and 11, and 25% in Year 10, fall into this category.

- *The only difference in class names is capitalisation.* In virtually all such cases,<sup>16</sup> the overlap between the membership of the two classes was over 75%. For example, looking at pupil 3, the overlap between the membership of 'Ma1' and 'ma1' was 96%. For pupil 4, the overlap between 'Ma4' and 'ma4' was 100%.

The measure of incidence reported in both Tables 4 and 6 is calculated once duplicates are removed.

Table 7: Examples of moves between classes within Year 11

Pupil	School	Year	Class code	Date registered in class	Class size
3	C	9	9ab/Ma1	18/03/2018	24
3	C	9	9ab/ma1	16/05/2018	25
4	D	9	9b/Ma4	18/03/2018	23
4	D	9	9b/Ma4	30/03/2018	23
4	D	9	9b/ma4	07/04/2018	23

### 2.5.2.2 Changes within seven days

Another distinct set of scenarios identifiable in the data are very short-period apparent memberships: a pupil is registered in a second class within seven days of being assigned to the first class. It is likely that this category includes different scenarios such as registration to different modules that are taught at the same time, actual class moves due to timetabling issues, or incorrect initial assignment of pupils, and so forth.

We do not consider any of these scenarios to be actual moves (Table 2 above does not record these as moves—they are excluded). Class membership in a trial is important inasmuch as it influences the mutual experience of the pupils in a group. However, if a group membership is short (measured in days), it is unlikely to have any notable effect on the pupil leaving the group, or on the group the pupil is leaving. While we do not consider these instances to be moves, they are nonetheless practically important for correctly identifying group membership in a trial.

### 2.5.2.3 Complex cases

The 'complex cases' category includes instances where the class code does not conform to a set form, or different code types are used in the same school at the same time, or the entry is affected by human error. All in all, this means that without an assumption-based manual process, with different rules used for different schools, we cannot ascertain whether these records suggest moves between classes, are duplicates or capture the pupils membership in several modules, or are coding errors.

We use this group to set the upper limit of the possible incidence of moves (displayed in Table 6); however, based on the manual assessment of some cases included in this category, the true incidence is within the range but without it reaching the upper bound.

It should also be noted that the number of complex records is a lot higher for Year 11. This could be explained by the use of revision classes whereby pupils are moved to a revision class for a period and then moved back to their initial class.

---

<sup>16</sup> There were 44 records in Year 9 and 20 records in Year 10 where the overlap between the classes identified by the lower- versus upper-case names was below 75%.

### 3. Between-year changes to class membership

#### 3.1 Revisiting the definition of classes and moves

The idea of moving between years is slightly more complex: often classes are wholly reconfigured with changes to both their name as well as composition. The concept of a 'move' as defined for the within-year case might not be useful when trying to understand changes between academic years. Instead, we argue it is more accurate and directly relevant to the aim of this project to first and foremost think of changes to class composition instead of moves.

To correctly identify changes to class composition between years, once again we need to consider how we define the concept of 'class'. In addition to the nominal definition that suggests that a class is defined by its name, when tracking pupils across years we also need to take into account the composition of the class. We posit that:

*A class is defined by its membership, it is nothing more, nothing less, than the group of pupils that study in the same group (and potentially with the same teacher) over a period of time.*

Under this definition, if a predetermined, minimum number of pupils leave the group, or join it, the class is no longer the same. The literature recognises that pupils are nested in classes and, thus, students in a certain classroom share experiences that are not shared by students in other classes (Black and Wiliam, 2005).

Defining the class as the group of pupils who study together (with the same teacher) at a given point in time means that to track changes to the composition of a class we need to assess the overlap in terms of student membership between classes in successive years.

If that overlap is 100%, we know for certain that we are seeing the same class across years. If the overlap is above a high threshold<sup>17</sup> but below 100%, we argue that the class itself is the same: they are mostly the same pupils that form the same group, but with some pupils changing classes. Finally, if the overlap between classes is below the threshold, we argue that the groups being compared are different classes (as their memberships are substantially different).<sup>18</sup>

As such, we define a move as an instance in which large changes occur to the membership of the class the pupil is a member of. We define a change to be 'large' when the overlap between the membership of the two classes (lower versus higher year) drops below a set threshold.

However, there is a complication. Classes cannot be directly identified through their membership; we need to use their 'names' (their codes) to identify them. This raises important additional questions. If we define a class to be the same across years when the overlap of pupils between the two years is high, is this reflected in the similarity of their name? Do codes consistently identify classes across years (for example, 9S/Ma1 and 10S/Ma1)? Alternatively, are there situations when the class names change substantially but the membership does not,<sup>19</sup> or instances when the name stays identifiably the same but the membership changes?

To understand the interaction between class membership and class name across years we build and test four scenarios that compare classes across two years:

- Scenario 1: high membership overlap and same name;
- Scenario 2: high membership overlap and different name;
- Scenario 3: low membership overlap and same name; and
- Scenario 4: low membership overlap and different name.

---

<sup>17</sup> The question of what a high overlap is, is crucial and we discuss it further on. We argue for a cut-off at 75%.

<sup>18</sup> We calculate the overlap with reference to the membership of the lower year class, in this way we track if the pupils move together into a class upper year.

<sup>19</sup> Class 'names' can be linked to a particular teacher. So, if the teacher changes between years, then the class name might change even though the membership of the class might remain broadly the same.

Scenarios 3 and 4 taken together can be used to quantify moves, while the remaining scenarios can shed light on researchers' ability to identify changes in classes by using class names. This issue has a crucial practical implication that transcends this research: when working with classes, as education researchers, can we correctly identify them without having to access their detailed memberships across years?

### 3.2 Prevalence of between-year moves

The analysis undertaken on changes to class membership between academic years suggests that the phenomenon is extensive.

Table 8: Between-year class moves

Changes between academic years	Proportion of pupils in the most recent year in classes with a substantially different composition <sup>20</sup>
Year 9 to Year 10	57%
Year 10 to Year 11	31%

Our results (Tables 9 and 10) show that between Years 9 and 10 over half of pupils from Year 9 took up places in classes where the overlap of membership with their previous class was below 75%. When looking at changes between Years 10 and 11, we find that approximately 30% of pupils move to classes that are substantially different in terms of membership.

The high incidence of moves between Years 9 and 10 compared to those between Years 10 and 11 is perhaps unsurprising given that Year 9 concludes Key Stage 3 and Years 10 and 11 correspond to Key Stage 4 and conclude with GCSE examinations.

Table 9: Between-year changes in class membership

Changes between academic years	Overlap criteria	Total pupils in classes with membership overlap	Total pupils in classes without membership overlap
Year 9 to Year 10	75%	7,638	11,789
Year 10 to Year 11		12,863	7,062

Table 10: Between-year changes in class membership (75% overlap)

Changes between academic years	Total pupils in both years	Total pupils in classes without membership overlap	Moves as a proportion of pupils in the lower year
Year 9 to Year 10	20,168	11,789	57%
Year 10 to Year 11	20,652	7,062	34%

<sup>20</sup> Threshold set at 75%. This means that for a pupil to have been deemed to have moved class, the class in which they are found in the most recent school year must have a composition in terms of other pupils which changes by more than 25%.

### 3.3 Pupil- and school-level characteristics associated with between-year changes

#### 3.3.1 School characteristics

We implemented regression analysis to determine whether there is any association between change in class composition between years and school characteristics. A statistically significant association between changing composition and school size was found. As school size increases, so does the number of changes.

#### 3.3.2 Pupil characteristics

We used mixed-effects logistic regression analysis as well as fixed-effects logistic regressions to assess if there are links between pupil characteristics and the pattern of their class membership across years.

The results show no significant differences when we tested class changes based on differences in class composition (irrespective of class name).

However, when we tested the link between pupil characteristics and class changes, with classes defined by name, we found that pupils who were eligible for free school meals were more likely to change classes between years, that is, they were more likely to be in a class of a different name. The results hold irrespective of whether the class they move to is composed by the same pupils or not.

### 3.4 Relevance to EEF trials

Between-year moves are relevant to EEF studies spanning at least two school years (21% of trials according to a recent review)<sup>21</sup> and more specifically studies comprising students at Key Stage 4 where ability grouping is the norm (5% of trials).<sup>21</sup>

Our view is that for the primary analysis, class membership should be recorded at the commencement of the study. However, it would be informative to collect class membership data at the beginning of each new school year, and to perform sensitivity analysis to understand how changes to class membership affect the effect of the intervention on outcomes. Doing so requires the collection and analysis of class membership data, which is, admittedly, costly and burdensome, but could yield more precise estimates.

### 3.5 Data processing and determination of the prevalence of moves

#### 3.5.1 Class changes between Year 9 and Year 10

Table 11 illustrates the results when combining class names and class membership for moves between Year 9 and Year 10. Following a 75% threshold for membership overlap, we see two clear-cut scenarios: 46% of pupils from Year 9 move to classes that are intrinsically different based on their composition *and* have a different name, while 31% of pupils from Year 9 move to classes that have the same name as well as similar composition. Given the change from Key Stage 3 to Key Stage 4, it is unsurprising that classes are reconfigured and more pupils join 'new' classes.

More problematic, however, are the remaining two categories that harbour non-trivial contingents of pupils:

- 7% of pupils from Year 9 remain part of the same class by virtue of composition, but the name of the class changes; and
- conversely, 12% of pupils from Year 9 are part of classes in Year 10 that have the same name as the class they were in in Year 9, but the composition of the class changes.

---

<sup>21</sup><https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-evaluation-reports-and-research-papers/syntheses-of-eef-evaluations/review-of-eef-projects>

The two categories are important to bear in mind when planning and designing research that requires the ability to identify classes across years (Years 9 and 10, in this case). The analysis suggests that solely using the class name as a proxy for changes between years will lead to errors for approximately 20% of pupils.

Our result is relatively stable irrespective of how the overlap threshold is defined. Up until this point when discussing results, we did so based on a set threshold for defining ‘high’ overlap, arbitrarily set at 75%. Table 8 includes the results for other definitions of ‘high’ overlap that range between 60% and 90%.<sup>22</sup>

Table 11: Changes between Years 9 and 10 at different levels of overlap

Changes between academic years	Overlap criteria	Membership overlap and different class names	Membership overlap and consistent class names	Lack of membership overlap and different class names	Lack of membership overlap and consistent class names
Year 9 to Year 10	60%	11%	35%	42%	8%
	65%	10%	34%	44%	10%
	70%	8%	32%	45%	11%
	<b>75%</b>	<b>7%</b>	<b>31%</b>	<b>46%</b>	<b>12%</b>
	80%	6%	28%	47%	15%
	85%	4%	23%	49%	20%
	90%	3%	17%	50%	26%

Table 12: Changes between Years 10 and 11 at different levels of overlap

Changes between academic years	Overlap criteria	Membership overlap and different class names	Membership overlap and consistent class names	Lack of membership overlap and different class names	Lack of membership overlap and consistent class names
Year 10 to Year 11	60%	14%	59%	19%	5%
	65%	13%	57%	20%	7%
	70%	12%	54%	21%	10%
	<b>75%</b>	<b>11%</b>	<b>51%</b>	<b>22%</b>	<b>12%</b>
	80%	10%	45%	23%	18%
	85%	7%	37%	25%	26%
	90%	5%	26%	27%	38%

### 3.5.2 Class changes between Year 10 and Year 11

Table 12 contains the results for the changes between Year 10 and Year 11. In contrast with moves between Year 9 and Year 10, most pupils (51%) remain in the same class (both by name and composition) and 22% change classes as identified by both changes in membership and class name.

<sup>22</sup> While the 85% and 90% scenarios are included for completeness, we argue against their use as reasonable thresholds. Due to timetabling issues and other exogenous factors some flexibility is needed when defining overlap.

Similarly to the changes between Years 9 and 10, in moving between Years 10 and 11 we also find just over 20% of pupils would be misclassified by looking solely at class names:

- 11% of pupils in Year 10 remain in the same class based on membership, but the name of the class changes;
- conversely, 12% of pupils in Year 10 are in classes in Year 11 that have the same name as in Year 10, but have a different composition.

### 3.5.3 Identifying class names across years

The previous sections highlight the importance of class name in identifying pupil moves. Ignoring the question of class membership, Tables 11 and 12 show that 53% of pupils move to a class identified by a different name when moving from Year 9 to Year 10. This figure is 33% when looking at moves between Years 10 and 11.

Identifying whether a class name is the same between two years is not trivial and in analysing the data we identified several scenarios:

1. the coding structure of the class name is the same across years, which makes it easy to identify if the class name changes;
2. the coding structure is not similar but the two structures are comparable without having to make assumptions as to the meaning of the new elements of this structure;
3. the coding structures are comparable and a simple assumption can be used to pair classes;
4. the coding structures are different; and
5. the name is entirely different on account of class structures changing (for example, classes are combined or split).

Table 13: Examples of differences in naming classes across years

Scenario		Lower year class name	Upper year class name	Description
1	Same structure	9bs/Ma3	10bs/Ma3	Excluding the year indicator, the code is the same.
2	Comparable structures	9pr/Ma3	10pq/Ma3	In Year 10, code 'pr' is replaced by the code 'pq'. This is consistently done across all classes.
3	Comparable structures, some assumption required	9E/Ma2a	10X/Ma2a	The assumption here relates to the pairing of 'E' with 'X', 'K' with 'V' and 'S' with 'Z'. We were able to do that based on overlap being 0% on all other combinations.
		9K/Ma1b	10V/Ma1b	
		9S/Ma3b	10Z/Ma3b	
4	Different structures	9mq/ma 9mj/ma 9ml/ma 9mr/ma 9mm/ma 9mk/ma	10mc/ma 10gmb/ma 10rmb/ma 10gmc/ma 10bmb/ma 10bma/ma	The two columns list the names of the classes in the two years. The classes have different naming structures.
5	Different names due to restricting of classes	MAT9X-1-ABC	MAT10X-1-ACV	Classes have different naming structures because the two types of classes 'ABC' and 'ABK' were merged into 'ACV'.
		MAT9Y-1-ABK		

Table 13 includes examples of each of these scenarios. Following a conservative approach, in generating the results of this analysis we considered that the name is the same in scenarios 1, 2, and 3. In scenarios 4 and 5 we concluded that the class name changed. In practice, a researcher might use their judgement and might be able to pair some classes we categorise in scenario 4. However, in most cases, substantive input from schools would be needed to untangle class pairing from scenarios 4 and 5. In this research we were unable to match classes by name for 29 schools (across Years 9 and 10) and 19 schools (across Years 10 and 11).

Finally, it is also worth noting that the incidence of changes between Years 10 and 11 exceeded our expectations. The manual examination of the cases in which that structure of class codes did not match between the two years leads to an interesting conclusion. There are schools in which classes are wholly restructured between the two years. In some instances, classes from Year 10 are split into smaller classes in Year 11; in other instances, classes from Year 10 are combined in fewer classes in Year 11.

## 4. Recommendations

### 4.1 EEF evaluators and researchers interested in class membership

In thinking about the implications of these findings for trial design, it is worth first pausing to remember why analyses of pupil-level outcomes from cluster randomised trials require the inclusion of random effects at multiple levels. As with any social grouping or organisation, classes and schools shape individual behaviours and attitudes. Likewise, class teachers and headteachers often deliver programmes and interventions in noticeably different ways. This violates the assumption that participants in a trial are statistically independent.

These levels should be taken into account in regression models (Demack, 2019). This is particularly important in cluster randomised trials where schools are the unit of allocation but pupils the unit of analysis. We recommend fitting three-level hierarchical models, that is, pupils within classes and classes within schools. This recommendation applies particularly to evaluations of programmes at Key Stage 4 where ability grouping is common. However, the benefit of controlling for class membership needs to be weighed against the additional cost, burden and risks of such an approach.

Given this background, our findings give some cause for optimism and some obvious motivation for further research. First, if researchers are conducting a trial that involves following pupils over one school year, the problem of class to class moves is likely to be trivial and can be ignored. We do not mean by this that a researcher should not seek to account for class as well as the school level in their statistical work, just that it is adequate for researchers to take account of class membership at randomisation assuming that randomisation occurs at the beginning of the school year and need not concern themselves with class movements beyond this. As an aside, where the primary outcome is some composite measure of attainment such as an aggregate of English and maths attainment, or 'Attainment 8' measure, it is acceptable to ignore the class level altogether and account for clustering solely at the level of the school.

More concerning is the case where an evaluation requires following pupils over more than one school year. Our evidence shows that there is likely to be appreciable movement between classes over a two-year period such that a pupil's class at randomisation might not be the class in which they spend the majority of their time overall or be the class in which they receive instruction in the run-up to sitting examinations. Trials using GCSE results as primary outcome are prime candidates for the use of three-level hierarchical models.

The first challenge in this situation is collecting accurate data on class membership on a frequent enough basis. We realise that in many cases this will not be a practical or realistic proposition and researchers will resort to accounting for the clustering of observations at the school level only. There may be circumstances where it is possible to collect data on school membership at the beginning of the first school year over which pupils are followed and then again at the commencement of the second school year. Were such circumstances to materialise we would recommend sensitivity-testing results: first by accounting for class membership at the beginning of the first year at randomisation in the primary analysis of Key Stage 4 attainment data, and then re-running exactly the same models but accounting for class membership at the beginning of the second year.

### 4.2 Further research

Our view is that the findings presented in this report point to the need for further research on this topic. A strategy that may help shed some further light on the choices to be made would be to conduct some simulations. A number of simulated class to class moves at the class level could be incorporated into a general simulation of a stylised cluster randomised trial, where true population effect size is incorporated (for example  $ES = 0.20$ ) and drawing on real data

from a Key Stage 4 study to generate distributions of outcome variables and covariates that are realistic and which conform to the assumptions of normality. This data could then be analysed (1) ignoring class level completely, (2) accounting for class at randomisation, (3) accounting for class at some later point, and (4) accounting explicitly for class to class moves. The results from these four analyses can then be compared to the known population values to assess how these four strategies perform in terms of bias and precision.

### 4.3 School data managers

School data managers across England are encouraged to harmonise and document their class codes. This information could be included in their NPD and school census returns.

## Appendix A: Focused literature review summary

Educational research was one of the first fields in the social sciences to adopt and fully develop multilevel modelling due to the widespread presence of hierarchical data structures with two or more levels of analysis such as pupils, classrooms, or schools. The use of multilevel models is related, *inter alia*, to the type of research questions relevant to educational sciences and to the need to understand which is the most relevant analytical unit for measuring scholastic attainment. For instance, researchers and policy-makers want to know whether an intervention should focus on the class, the school, or on the students and how these different levels interact together in explaining scholastic performances.

One of the fields within educational sciences where multilevel modelling can be fruitfully applied is the investigation of how particular educational features and dynamics at classroom level affect student achievement, behaviours, and choices (Martínez 2012). Unfortunately, despite the importance of the classroom as an educational device, current research on the UK secondary education often fails to take into account that students are not completely independent. Rather, students are nested in classrooms and, thus, students in a certain classroom share experiences that are not shared by students in other classrooms (Black and Wiliam 2005). This review aims to uncover the importance of the classroom as a relevant analytical unit in the UK secondary education systems and to raise awareness on the necessity of using multilevel modelling when designing an intervention and analyse hierarchical educational data.

First, we review five prominent fields of research that focus on the impact of classroom characteristics on pupil achievement. Second, we emphasize the importance of the classroom as an educational device showing that, despite the classroom often being omitted from statistical analyses, a great amount of variation in students' outcomes is located at the classroom level. Third, we propose the use of multilevel modelling to take into account the nesting of students in different classrooms. In doing so, we outline how multilevel modelling can improve current research practices in educational research.

### The importance of the classroom in the UK secondary education

Similarly to other European curriculums, the education system in the UK is split into four different 'Key Stages'. Secondary education involves pupils from age 11 to 16 that ought to complete Key Stages 3 and 4. The most important assessment occurs at age 16 when students pursue their General Certificate of Secondary Education (GCSE). After the age of 16, education is optional. Educational researchers are usually interested in assessing the importance of classroom characteristics as an educational feature, in relation to class-room atmosphere, peer relationships, and academic achievements, especially GCSE test scores. In this review, we critically examine recent classroom-level literature with the aim of identifying the shortcomings of the current approaches in studying classroom dynamics. Specifically, we focus on five streams of research that are commonly investigated in relation to the UK secondary education system.

A first stream of research investigates the effects of class size on classroom processes, interactions and the learning environment (for a general review, see Peter Blatchford 2012). The UK has the fourth largest classrooms in the OECD with an average of 20 pupils in the secondary phase (Peter Blatchford and Webster 2018). Most attention in educational research has been paid to whether or not smaller classes lead to better academic outcomes for pupils by influencing classroom processes. In line with cross-national research (for example, Slavin 2010), the research on the topic presents a good deal of controversy over the effect of the class size on educational achievement. Some authors suggest that class size can have important implications for pupil learning (for example, Peter Blatchford, Bassett, and Brown 2011) while others argue that the effects of class size on teaching and learning are minimal (Levačić and Vignoles 2002). The lack of consensus is related to the fact that statistical analyses in the secondary phase often disregard classroom characteristics (e.g, Peter Blatchford, Bassett, and Brown 2011; Peter Blatchford and Webster 2018). For instance, in their systematic review on the topic, Blatchford (2003) state that in the UK 'there is little research to guide debates on class size effects, and such research as does exist is limited in terms of research methods' (2003, 570). A second prominent field of research is the benefits or disadvantages of different within-class ability grouping practices (for an extensive review see, Kutnick et al. 2005). In the UK, class-level grouping can be organised by 'setting' or 'streaming' (sometimes also called 'tracking'). The former refers to ability grouping for particular subjects. The latter is when ability grouping is for all subjects. Regardless of the type of grouping, the assumption behind ability grouping is that students

achieve better academic outcomes when the range of attainments in a class is homogeneous. However, this assumption rarely withstands empirical scrutiny. While homogeneity in attainment seems to have some benefit for some higher-attaining pupils (for example, Baines, Blatchford, and Kutnick 2003), lower-attaining pupils are often severely disadvantaged (Francis et al. 2017). Furthermore, most UK studies tend to focus exclusively on junior classrooms and not on secondary education (Key Stage 3 and 4) and thus fail to 'consider grouping contexts in relation to children's development or consider how teachers treat children differently at different ages.' (Baines, Blatchford, and Kutnick 2003, 12).

Third, a related topic in the study of classroom dynamics is the influence that the group exerts on the achievement of each student within it. In addition to ability grouping, other types of grouping are likely to influence students' performances, both throughout the year(s) and in terms in GCSE scores. For instance, the gender composition of the classroom is likely to affect pupils' achievements beyond the effect of their own gender. The effect of classroom gender composition recently gained interest due to high levels of participation and better academic achievement of girls in secondary education (editor 2020; Broecke and Hamed 2008). Academic research on the topic has mostly focused on the negative impact of the gender gap in academic achievement and aspirations within mixed-sex schools and single-sex schools (for example, Stables 1990). However, recent studies have started to take into account the classroom level comparing single-sex classes with coeducational classes within coeducational schools (for example, all-girls classes for mathematics) (for a literature review see, Belfi et al. 2012; Bramley, Rodeiro, and Vitello 2015). Unfortunately, the limited literature on the topic presents mixed evidence. Some studies show a positive effect of single-sex classrooms while others found an insignificant effect of gender compositions on classroom dynamics and student achievements (for example, Sullivan 2009; Kessels and Hannover 2008).

Two other grouping-related questions usually explored in sociology of education are worth mentioning. The first one is the relationship between socio-economic status and academic achievement (for a meta-review see Sirin 2005). The other is the increasing presence of black and ethnic minority students relative to white British students. Although both factors influence class composition and, thus, are likely to impact on students' performances through peer influence, most of the studies simply focus on students' characteristics and disregard the classroom level and class composition (for example, Battle and Lewis 2002).

Fourth, a fast-growing body of work explores the short- and long-term impact of changes to the UK National Curriculum on foreign language teaching (McLelland 2017). This body of research has mostly focused on student-level predictors such as enjoyment of the subject and students' self-efficacy and the effects of policy implementation at the school level (for a review see Lanvers 2017). However, some recent studies have also investigated how classroom practices regarding tests and examinations can have an impact on student engagement and performances. These studies found evidence that classroom activities and practices exert a significant impact on learning outcomes (Wingate 2018; Graham et al. 2017; Mitchell and Myles 2019).

A final and more recent stream of research deals with the concentration of students with special educational needs and disabilities (SEND). In 2016, approximately 2.8% of the pupil population in England had a SEND (Blatchford and Webster 2018). There has been interest in the effect of classroom processes and teaching on the education of pupils with and without SEND. In the UK, particular attention has been paid to the effect of class-level 'setting' versus 'streaming' (for example, Peter Blatchford and Webster 2018), the interaction between teachers and pupils (for example, Bunning and Ellis 2010), or other classroom dynamics (Webster and Blatchford 2015). However, despite the increase of children with SEND over the last few decades (for example see Thomas and Vaughan 2004), the scarcity of research on the topic impedes our understanding of how including pupils with SEND in mainstream settings might impact classroom dynamics (Blatchford and Webster 2018).

## Does the classroom matter?

Two conclusions can be drawn from our literature review. First, in line with Bronfenbrenner (1979) model, the reviewed studies reveal the effect of classroom composition on individual behaviours. As Bronfenbrenner as argued, learning happens within concentric 'microsystems'. Core systems, especially the classroom, have forces that are different from peripheral systems (for example, schools, neighbourhoods) (Kounin and Gump 1974). Second, in addition to its effect on learning experience, classroom composition explains some of the variance in terms on achievement. Opdenakker and Damme (2000) found that 23% of the variation in students' achievement at the end of the first grade of secondary school could be attributed to class-level variables. Similarly, Demack (2019) found that between 30% and 70% of the

variation in students' achievement in two recent clustered RTCs can be attributed to the classroom in the English secondary phase.

Despite the relevance of the classroom dynamics and characteristics, the classroom is often omitted from statistical analyses looking at the UK secondary education system. In their systematic review of RTCs in education research, Connolly, Keenan, and Urbanska (2018) found that out of the 40 RCTs conducted in English secondary schools, only half directly acknowledge the clustered nature of the data within the trial design. Observational studies show a similar trend. In their literature review, Levačić and Vignoles (2002) found that the classroom level is often omitted from the statistical analysis.

## Why the classroom is often disregarded

Our previous points raise the question of why researchers do not include the classroom level in their analyses. In many disciplines, it is common to randomise individuals rather than groups. For instance, in a medical setting some participants receive—at random—the trial drug (treatment) and some individuals receive a placebo (control). In education research, however, this type of experimental design is not always possible. Instead, a common experimental protocol is to measure student outcomes when an intervention is administered to a limited number of different classes. In this case, the intervention (treatment) is randomised by class, not by student. Yet, the outcome of interest is often measured at the student level (that is, students' performance) and not at the class level since only a few classes have been assigned to the trial (Theobald 2018).

Concerning observational data, a common statistical approach used to analyse educational data is ordinary least squares (OLS) regression analysis. For example, OLS regression can be used to understand what classroom characteristics impact students' scholastic performance, controlling for class-level and other variables. Typically, a standard OLS model might try to model hierarchical data structures using what is commonly referred to as disaggregated variables. A disaggregated variable is created by assigning to each student in a specific group the same value for the group characteristic (for example, students enrolled in single-sex versus coeducational sections).

However, in both cases, modelling hierarchical data while not taking into account their complex nested structure may produce inaccurate results about the statistical significance of a relationship. For instance, a standard OLS model that uses aggregated variables will use the individual-level sample size, which is inaccurate for an aggregate variable measured at the group level. This practice can also lead to incorrect conclusions about the significance of the relationships under investigation leading to meaningless tests of significance and an inflated risk of Type I error. This can be problematic for theory testing since bias results can serve as false statistical evidence against valid theories and past studies.

## Conclusions

The main conclusion to be drawn from this review is that evaluations of programmes implemented in English secondary schools would benefit from a 'multisystemic' approach, whereby both school-level and class-level variables are accounted for in impact analyses.

## Appendix B: Technical Note

The paragraphs below include the details of the methodology used to carry out the analysis, including relevant decisions made.

### Data preparation

The data used in the analysis was extracted from a dataset that contains pupils' class memberships in all subjects in 146 schools. The dataset contains 1,014,994 records.

As a first step the data was cleaned and readied for analysis. We performed several actions:

1. Diagnose and assess class subjects. The records for some classes do not contain information on the class subject.
  - 75% of records, pertaining to 143 schools, contained information of class subject;
  - 25% of records, pertaining to 144 schools, did not contain information on the class subject.
2. Adjust class subject descriptions: For records with information on the class subject, we assessed the descriptions of subjects across schools to ensure consistency. Out of the 75% of records with information on class subject 9.8% were maths classes and a further 10.4% were English classes.
3. Derive class subject: For cases in which the class subject was not available, we derived the class subject by using the information contained in the class codes. We only worked to identify maths and English classes. This was done by assessing the correspondence between the class codes and subjects for classes where both these items of information existed. As such, codes 'Ma' were taken to refer maths and 'En' to English. Out of the 25% of records without information on class subject, approximately 6.2% were identified to be maths classes and a further 5.8% English classes.
4. Remove unusually-sized classes: classes that contained fewer than 10 pupils or more than 50 pupils were removed from the analysis.
  - Classes with fewer than 10 pupils: just over 21,000 records were removed;
  - Classes with more than 50 pupils: just over 7,000 records were removed.
5. The data was truncated to include only those records that were relevant for this research. Two analytical datasets were created:
  - The main analytical dataset containing the records pertaining to membership in maths classes;
  - A secondary dataset that includes membership in both English and maths to be used to determine the level of grouping used by the school.

### Developing a measure of grouping

Using the secondary analytical dataset, we set about to develop a measure of grouping by examining the class compositions between and within subject areas in a school (that is, maths, English). Schools can use one of two forms of grouping: streaming or setting. Streaming occurs when a school places pupils in prior attainment-based groups across subjects. As such, it is likely that class composition will be relatively consistent across subjects in a stream. Setting occurs when a school places pupils in performance-based classes by subject-specific prior attainment. However, a single school can use both forms of grouping across different subjects. For example, setting can be used for maths and English, while streaming could be used for other subjects.

Our analysis showed that it is feasible to identify if a school uses grouping in assigning pupils to classes. However, using solely data on class composition, we do not think it feasible to clearly determine whether it is setting or streaming that is being used. In addition to pupils' attainment, schools also take into account other factors when assigning pupils to

classes, such as pupil behaviour or logistical issues (timetabling constraints, room sizes, etc). As such, we developed a continuous measure that quantifies the extent to which a school groups pupils across classes, without specifying what type of grouping it is.

To this end, we calculated the proportion of overlap of pupils between each pairwise combination of classes across main subjects (maths, English). This can be defined as the conditional probability of a pupil being in a class in maths, given their membership of a class in a different subject, such as English.

Using the results of this analysis we developed two complementary measures:

1. The average of the probabilities across all pairwise combinations of classes across maths and English. This was accompanied by the coefficient of variation (ratio of the standard deviation to the mean) to quantify variability.
2. The proportion of pairwise combination of classes where the overlap is above a high threshold of 70%. For simplicity, the latter variable was used in the analyses of the link between grouping and moves.

## Assessment of within-year moves

The assessment of within-year moves was carried out using the main analytical dataset in the following steps:

- Separately for each academic year we identified the pupils with more than one record;
- Eliminated any inconsistent dates (records that fall outside the study period);
- Identify and remove records that we term as being 'duplicates' (classes with the same name, once capitalisation is removed). As a back-up we also check that the overlap between such classes is over 90%.
- Identify and remove records that are registered within 7 days of the initial record;
- Identify and remove moves by assessing the differences between the class codes;
- The remaining records are complex cases that cannot be classified without additional information.

Once all records were classified, two analytical datasets were developed and used in further analyses focusing on schools and pupils.

1. We developed an aggregate school-level dataset that indicates the number of records within each of the categories outlined above for each school and for each year. Additional data was then added into this dataset:
  - Trial arm: treatment or control;
  - Our measure of pupil grouping;
  - School characteristics, including: school type; school size, % girls; % SEN; % EAL; % FSM; pupil-teacher ratio; previous KS4 results (% grades 9-4 in English and maths).
2. We also developed a pupil-level dataset that contains a series of dichotomous variables to indicate what categories each pupil is present in. The dataset was then enriched with the school-level characteristics outlined above. In addition, we included pupil-level information: gender, month of birth and FSM status.

## Inferential analyses of within-year moves

### School-level analysis

Table A1: The effect of school-level characteristics on the number of moves per school

	<b>Model 1: Year 10</b>	<b>Model 2: Year 11</b>
<i>Predictors</i>	<i>OLS regression</i>	<i>OLS regression</i>
Intercept	46.33 * (6.42 – 86.25)	-4.17 (-36.14 – 27.80)
Treatment school (ref. control)	5.58 * (0.07 – 11.08)	-0.35 (-4.76 – 4.06)
School size	-0.00 (-0.01 – 0.01)	0.01 * (0.00 – 0.01)
% girls	-0.10 (-0.31 – 0.11)	-0.05 (-0.21 – 0.12)
% SEN	-0.59 * (-1.11 – -0.06)	0.10 (-0.32 – 0.53)
% EAL	0.03 (-0.15 – 0.21)	0.00 (-0.14 – 0.15)
% FSM	0.54 * (0.00 – 1.08)	0.22 (-0.22 – 0.65)
Teacher-pupil ratio	-1.36 (-3.15 – 0.43)	-0.02 (-1.46 – 1.41)
Not an academy (ref. academy)	-5.36 (-11.66 – 0.94)	2.14 (-2.90 – 7.19)
Attainment: % grades 9-4 in English and maths	-0.18 (-0.48 – 0.13)	0.04 (-0.21 – 0.28)
Ability grouping (ref. mixed-ability)	-0.15 (-0.33 – 0.02)	-0.04 (-0.19 – 0.10)
Observations	122	122
R <sup>2</sup> / R <sup>2</sup> adjusted	0.219 / 0.148	0.074 / -0.009

\*  $p < 0.05$  \*\*  $p < 0.01$  \*\*\*  $p < 0.001$

## Pupil-level analysis

Table A2: The effect of pupil and school characteristics on moving classes

	<b>Model 1: Year 10</b>	<b>Model 2: Year 11</b>
<i>Predictors</i>	<i>Mixed-effects logistic</i>	<i>Mixed-effects logistic</i>
Intercept	0.10 (0.00 – 314.90)	0.00 (0.00 – 3.94)
Gender: Girls (ref. boys)	0.94 (0.75 – 1.18)	0.61 ** (0.45 – 0.82)
Month of birth: Feb (ref. Jan)	0.86 (0.52 – 1.42)	1.02 (0.50 – 2.06)
Month of birth: March (ref. Jan)	0.59 (0.34 – 1.02)	1.17 (0.60 – 2.26)
Month of birth: April (ref. Jan)	1.01 (0.60 – 1.68)	0.67 (0.31 – 1.43)
Month of birth: May (ref. Jan)	0.97 (0.59 – 1.58)	1.07 (0.54 – 2.09)
Month of birth: June (ref. Jan)	0.92 (0.55 – 1.53)	1.00 (0.50 – 2.01)
Month of birth: July (ref. Jan)	1.09 (0.67 – 1.77)	1.32 (0.69 – 2.53)
Month of birth: Aug (ref. Jan)	0.97 (0.59 – 1.58)	1.20 (0.63 – 2.31)
Month of birth: Sept (ref. Jan)	0.71 (0.42 – 1.18)	1.76 (0.95 – 3.27)
Month of birth: Oct (ref. Jan)	0.86 (0.53 – 1.42)	1.39 (0.74 – 2.62)
Month of birth: Nov (ref. Jan)	0.84 (0.51 – 1.39)	1.38 (0.72 – 2.65)
Month of birth: Dec (ref. Jan)	0.82 (0.49 – 1.37)	1.24 (0.64 – 2.40)
Eligible FSM 6 (ref. not eligible)	1.20 (0.95 – 1.52)	0.85 (0.62 – 1.17)
School size	1.00 (1.00 – 1.00)	1.00 (1.00 – 1.00)
% girls	1.00 (0.97 – 1.04)	0.97 (0.93 – 1.02)
% SEN	0.98 (0.89 – 1.08)	1.10 (0.98 – 1.24)
% EAL	0.98 (0.95 – 1.02)	0.98 (0.94 – 1.02)
% FSM	1.13 * (1.02 – 1.25)	1.05 (0.93 – 1.20)
Teacher-pupil ratio	0.82 (0.54 – 1.24)	1.33 (0.77 – 2.31)

Not an academy (ref. academy)	2.00 (0.61 – 6.49)	1.71 (0.37 – 7.89)
Attainment: % grades 9-4 in English and maths	1.03 (0.98 – 1.09)	1.01 (0.95 – 1.09)
Ability grouping (ref. mixed-ability)	0.97 (0.92 – 1.03)	0.99 (0.92 – 1.06)

#### Random Effects

$\sigma^2$	3.29	3.29
T00	2.31 SchoolURN	3.45 SchoolURN
ICC	0.41	0.51
N	46 SchoolURN	46 SchoolURN
Observations	6978	6978
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.181 / 0.519	0.119 / 0.570

\*  $p < 0.05$  \*\*  $p < 0.01$  \*\*\*  $p < 0.001$

## Assessment of between-year moves

Between-years moves were assessed in two separate analyses that correspond to the transition between Year 9 and Year 10, and Year 10 and Year 11, respectively. In each analysis, we followed the steps below:

- Truncated the dataset to ensure that the lower-year data set contained only the most recent records, while the upper-year dataset contained only the records closest to the start of the year. This was done to ensure that the analysis of between-year moves was not affected by within-year moves.
- Calculated the overlap between each class in the lower and upper year;
- Assessed if the form of the class codes matched between the lower and upper year. In instances where there was no automatic match, manual checks were performed and matches executed if the data supported it.

Records were classified into the four groups described in the report. Again, two analytical datasets were developed and used in further analyses.

1. An aggregate school-level dataset that indicates the numbers within each of the four scenarios was created. The same additional data was included as included in the within-year case.
2. A pupil-level dataset was also developed and, in addition to the variable identifying the scenario the pupil was in, also included the additional data as outlined for the within-year case.

## Inferential analyses of between-year moves

### School-level analysis

Table A3: The effect of school-level characteristics on the number of moves

	<b>Model 1: Year 10</b>	<b>Model 2: Year 11</b>
<i>Predictors</i>	<i>OLS regression</i>	<i>OLS regression</i>
Intercept	13.47 (-71.39 – 98.33)	-33.65 (-124.77 – 57.48)
Treatment school (ref. control)	-0.36 (-11.48 – 10.76)	5.18 (-7.07 – 17.42)
School size	0.03 *** (0.02 – 0.05)	0.02 * (0.00 – 0.04)
% girls	0.18 (-0.25 – 0.61)	-0.06 (-0.51 – 0.40)
% SEN	0.37 (-0.86 – 1.61)	0.83 (-0.56 – 2.21)
% EAL	-0.02 (-0.38 – 0.35)	0.08 (-0.34 – 0.51)
% FSM	-0.20 (-1.42 – 1.01)	0.99 (-0.31 – 2.29)
Teacher-pupil ratio	-0.70 (-4.28 – 2.87)	2.03 (-2.04 – 6.11)
Not an academy (ref. academy)	0.53 (-11.98 – 13.03)	-5.06 (-18.97 – 8.85)
Attainment: % grades 9-4 in English and maths	-0.26 (-0.90 – 0.37)	-0.21 (-0.92 – 0.50)
Ability grouping (ref. mixed-ability)	0.43 (-0.07 – 0.94)	-0.19 (-0.57 – 0.20)
Observations	106	114
R <sup>2</sup> / R <sup>2</sup> adjusted	0.166 / 0.078	0.174 / 0.094

\*  $p < 0.05$  \*\*  $p < 0.01$  \*\*\*  $p < 0.001$

## Pupil-level analysis

Table A4: The effect of pupil and school-level characteristics on moving classes

	<b>Model 1: Year 10</b>	<b>Model 2: Year 11</b>
<i>Predictors</i>	<i>Mixed-effects logistic</i>	<i>Mixed-effects logistic</i>
Intercept	0.01 (0.00 – 258.84)	0.01 (0.00 – 258.84)
Gender: Girls (ref. boys)	0.99 (0.85 – 1.14)	0.99 (0.85 – 1.14)
Month of birth: Feb (ref. Jan)	0.93 (0.66 – 1.33)	0.93 (0.66 – 1.33)
Month of birth: March (ref. Jan)	1.01 (0.72 – 1.42)	1.01 (0.72 – 1.42)
Month of birth: April (ref. Jan)	1.00 (0.70 – 1.43)	1.00 (0.70 – 1.43)
Month of birth: May (ref. Jan)	0.95 (0.67 – 1.34)	0.95 (0.67 – 1.34)
Month of birth: June (ref. Jan)	1.04 (0.74 – 1.47)	1.04 (0.74 – 1.47)
Month of birth: July (ref. Jan)	0.96 (0.69 – 1.33)	0.96 (0.69 – 1.33)
Month of birth: Aug (ref. Jan)	0.93 (0.66 – 1.32)	0.93 (0.66 – 1.32)
Month of birth: Sept (ref. Jan)	0.89 (0.64 – 1.26)	0.89 (0.64 – 1.26)
Month of birth: Oct (ref. Jan)	0.94 (0.66 – 1.33)	0.94 (0.66 – 1.33)
Month of birth: Nov (ref. Jan)	0.99 (0.70 – 1.40)	0.99 (0.70 – 1.40)
Month of birth: Dec (ref. Jan)	0.99 (0.69 – 1.40)	0.99 (0.69 – 1.40)
Eligible FSM 6 (ref not eligible)	1.09 (0.91 – 1.31)	1.09 (0.91 – 1.31)
School size	1.00 (1.00 – 1.00)	1.00 (1.00 – 1.00)
% girls	1.04 (0.99 – 1.08)	1.04 (0.99 – 1.08)
% SEN	0.96 (0.81 – 1.12)	0.96 (0.81 – 1.12)
% EAL	1.00 (0.96 – 1.04)	1.00 (0.96 – 1.04)
% FSM	1.02 (0.88 – 1.18)	1.02 (0.88 – 1.18)
Teacher-pupil ratio	1.02 (0.62 – 1.67)	1.02 (0.62 – 1.67)

Not an academy (ref. academy)	1.80 (0.42 – 7.68)	1.80 (0.42 – 7.68)
Attainment: % grades 9-4 in English and maths	0.99 (0.92 – 1.06)	0.99 (0.92 – 1.06)
Ability grouping (ref. mixed-ability)	1.04 (0.99 – 1.09)	1.04 (0.99 – 1.09)
<b>Random Effects</b>		
$\sigma^2$	3.29	3.29
T00	3.90 SchoolURN	3.90 SchoolURN
ICC	0.54	0.54
N	45 SchoolURN	45 SchoolURN
Observations	6840	6840
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.058 / 0.569	0.058 / 0.569
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$		

Table A5: The effect of pupil and school-level characteristics on moving to a class with a different name

	<b>Model 1: Year 10</b>	<b>Model 2: Year 11</b>
<i>Predictors</i>	<i>Mixed-effects logistic</i>	<i>Mixed-effects logistic</i>
Intercept	0.00 (0.00 – 171.03)	0.00 (0.00 – 1.88)
Gender: Girls (ref. boys)	1.08 (0.94 – 1.24)	0.88 (0.75 – 1.04)
Month of birth: Feb (ref. Jan)	1.22 (0.87 – 1.72)	1.19 (0.80 – 1.77)
Month of birth: March (ref. Jan)	0.86 (0.61 – 1.20)	1.20 (0.81 – 1.76)
Month of birth: April (ref. Jan)	0.91 (0.64 – 1.29)	1.19 (0.80 – 1.78)
Month of birth: May (ref. Jan)	0.89 (0.63 – 1.24)	1.22 (0.83 – 1.80)
Month of birth: June (ref. Jan)	1.02 (0.73 – 1.43)	1.03 (0.69 – 1.54)
Month of birth: July (ref. Jan)	1.03 (0.74 – 1.43)	1.21 (0.82 – 1.76)
Month of birth: Aug (ref. Jan)	1.17 (0.84 – 1.64)	1.34 (0.91 – 1.96)
Month of birth: Sept (ref. Jan)	0.86 (0.62 – 1.20)	1.12 (0.76 – 1.65)
Month of birth: Oct (ref. Jan)	0.86 (0.62 – 1.21)	1.47 * (1.00 – 2.15)
Month of birth: Nov (ref. Jan)	1.07 (0.77 – 1.50)	1.71 ** (1.17 – 2.50)
Month of birth: Dec (ref. Jan)	0.69 * (0.49 – 0.98)	1.15 (0.77 – 1.71)
Eligible FSM 6 (ref. not eligible)	1.36 *** (1.16 – 1.61)	1.34 ** (1.12 – 1.62)
School size	1.00 (1.00 – 1.00)	1.00 (1.00 – 1.00)
% girls	1.08 (0.99 – 1.18)	1.00 (0.96 – 1.04)
% SEN	0.82 (0.65 – 1.04)	1.08 (0.95 – 1.23)
% EAL	0.97 (0.91 – 1.04)	1.00 (0.96 – 1.04)
% FSM	1.10 (0.88 – 1.37)	1.09 (0.97 – 1.22)
Teacher-pupil ratio	1.11 (0.52 – 2.37)	1.34 (0.90 – 2.01)
Not an academy (ref. academy)	0.72 (0.07 – 7.23)	0.61 (0.17 – 2.11)

Attainment: % grades 9-4 in English and maths	1.10 (0.98 – 1.23)	0.97 (0.91 – 1.02)
Ability grouping (ref. mixed-ability)	1.06 (0.95 – 1.17)	1.04 * (1.01 – 1.08)

#### Random Effects

$\sigma^2$	3.29	3.29
T00	9.78 SchoolURN	2.30 SchoolURN
ICC	0.75	0.41
N	45 SchoolURN	41 SchoolURN
Observations	6840	5934
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.294 / 0.822	0.215 / 0.538

\*  $p < 0.05$  \*\*  $p < 0.01$  \*\*\*  $p < 0.001$

## References

- Baines, E., Blatchford, P. and Kutnick, P. (2003) 'Changes in Grouping Practices over Primary and Secondary School', *International Journal of Educational Research*, 39 (1), pp. 9–34. [https://doi.org/10.1016/S0883-0355\(03\)00071-5](https://doi.org/10.1016/S0883-0355(03)00071-5)
- Battle, J., and Lewis, M. (2002) 'The Increasing Significance of Class: The Relative Effects of Race and Socioeconomic Status on Academic Achievement', *Journal of Poverty*, 6 (2), pp. 21–35. [https://doi.org/10.1300/J134v06n02\\_02](https://doi.org/10.1300/J134v06n02_02)
- Belfi, B., Goos, M., de Fraine, B. and van Damme, J. (2012) 'The Effect of Class Composition by Gender and Ability on Secondary School Students' School Well-Being and Academic Self-Concept: A Literature Review', *Educational Research Review*, 7 (1), pp. 62–74. <https://doi.org/10.1016/j.edurev.2011.09.002>
- Black, P. and Wiliam, D. (2005) *Inside the Black Box: Raising Standards Through Classroom Assessment*, Granada Learning.
- Blatchford, P. (2003) 'A Systematic Observational Study of Teachers' and Pupils' Behaviour in Large and Small Classes', *Learning and Instruction*, 13 (6), pp. 569–95. [https://doi.org/10.1016/S0959-4752\(02\)00043-9](https://doi.org/10.1016/S0959-4752(02)00043-9)
- Blatchford, P. (2012) 'Three Generations of Research on Class-Size Effects', in *APA Educational Psychology Handbook, Vol 2: Individual Differences and Cultural and Contextual Factors*, pp. 530–54. APA Handbooks in Psychology. Washington, DC: American Psychological Association. <https://doi.org/10.1037/13274-021>
- Blatchford, P., Bassett, P. and Brown, P. (2011) 'Examining the Effect of Class Size on Classroom Engagement and Teacher–Pupil Interaction: Differences in Relation to Pupil Prior Attainment and Primary vs. Secondary Schools', *Learning and Instruction*, 21 (6), pp. 715–30. <https://doi.org/10.1016/j.learninstruc.2011.04.001>
- Blatchford, P. and Webster, R. (2018) 'Classroom Contexts for Learning at Primary and Secondary School: Class Size, Groupings, Interactions and Special Educational Needs', *British Educational Research Journal*, 44 (4). pp. 681–703. <https://doi.org/10.1002/berj.3454>
- Bramley, T., Rodeiro, C. V. and Vitello, S. (2015) 'Gender Differences in GCSE', Cambridge Assessment Research Report, Cambridge, UK: Cambridge Assessment, p. 44.
- Broecke, S. and Hamed, J. (2008) 'Gender Gaps in Higher Education Participation: An Analysis of the Relationship Between Prior Attainment and Young Participation by Gender, Socio-Economic Class and Ethnicity' (578942011-001), American Psychological Association. <https://doi.org/10.1037/e578942011-001>
- Bronfenbrenner, U. (1979) 'Contexts of Child Rearing: Problems and Prospects', *American Psychologist*, 34 (10), pp. 844–50. <https://doi.org/10.1037/0003-066X.34.10.844>
- Bunning, K. and Ellis, M. (2010) 'A Preliminary Investigation into Communication in the Special Needs Classroom During Key Stage 3 English Lessons', *Child Language Teaching and Therapy*, 26 (2), pp. 180–94. <https://doi.org/10.1177/0265659010368752>
- Connolly, P., Keenan, C. and Urbanska, K. (2018) 'The Trials of Evidence-Based Practice in Education: A Systematic Review of Randomised Controlled Trials in Education Research 1980', *Educational Research*, 60 (3), pp. 276–91. <https://doi.org/10.1080/00131881.2018.1493353>
- Demack, S. (2019) 'Does the Classroom Level Matter in the Design of Educational Trials? A Theoretical and Empirical Review', London: Education Endowment Foundation, EEF Research Paper, no. 3.
- Dumay, X. and Dupriez, V. (2008) 'Does the School Composition Effect Matter? Evidence from Belgian Data', *British Journal of Educational Studies*, 56 (4), pp. 440–77. <https://doi.org/10.1111/j.1467-8527.2008.00418.x>
- Adams, R. (2020) 'GCSE Results: Proportion of Entries with Highest Grades Soars.' *The Guardian*, 20 August 2020.

- Epple, D. and Romano, R. E. (1998) 'Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects', *American Economic Review*, 88 (1), pp. 33–62.
- EEF (2018) 'Statistical analysis guidance for EEF evaluations', London: Education Endowment Foundation.  
[https://educationendowmentfoundation.org.uk/public/files/Grantee\\_guide\\_and\\_EEF\\_policies/Evaluation/Writing\\_a\\_Protocol\\_or\\_SAP/EEF\\_statistical\\_analysis\\_guidance\\_2018.pdf](https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf)
- Francis, B., Archer, L., Hodgen, J., Pepper, D., Taylor, B. and Travers, M-C. (2017) 'Exploring the Relative Lack of Impact of Research on "Ability Grouping" in England: A Discourse Analytic Account', *Cambridge Journal of Education*, 47 (1), pp. 1–17. <https://doi.org/10.1080/0305764X.2015.1093095>
- Graham, S., Courtney, L., Marinis, T. and Tonkyn, A. (2017) 'Early Language Learning: The Impact of Teaching and Teacher Factors.' *Language Learning*, 67 (4), pp. 922–58. <https://doi.org/10.1111/lang.12251>
- Harker, R. and Tymms, P. (2004) 'The Effects of Student Composition on School Outcomes.' *School Effectiveness and School Improvement*, 15 (June). <https://doi.org/10.1076/sesi.15.2.177.30432>
- Kessels, U. and Hannover, B. (2008) 'When Being a Girl Matters Less: Accessibility of Gender-Related Self-Knowledge in Single-Sex and Coeducational Classes and Its Impact on Students' Physics-Related Self-Concept of Ability', *British Journal of Educational Psychology*, 78 (pt 2), pp. 273–89. <https://doi.org/10.1348/000709907X215938>
- Kounin, J. S. and Gump, P. V. (1974) 'Signal Systems of Lesson Settings and the Task-Related Behavior of Preschool Children', *Journal of Educational Psychology*, 66 (4), pp. 554–62. <https://doi.org/10.1037/h0036748>.
- Kutnick, P., Sebba, J., Blatchford, P., Galton, M. and Thorp, J. (2005) 'The Effects of Pupil Grouping: Literature Review', Nottingham : DfES.
- Lanvers, U. (2017) 'Contradictory Others and the Habitus of Languages: Surveying the L2 Motivation Landscape in the United Kingdom', *Modern Language Journal*, 101 (3), pp. 517–32. <https://doi.org/10.1111/modl.12410>
- Levačić, R. and Vignoles, A. (2002) 'Researching the Links Between School Resources and Student Outcomes in the UK: A Review of Issues and Evidence', *Education Economics*, 10 (3), pp. 313–31.  
<https://doi.org/10.1080/09645290210127534>
- Martínez, J. F. (2012) 'Consequences of Omitting the Classroom in Multilevel Models of Schooling: An Illustration Using Opportunity to Learn and Reading Achievement', *School Effectiveness and School Improvement*, 23 (3), pp. 305–26. <https://doi.org/10.1080/09243453.2012.678864>
- McLelland, N. (2017) *Teaching and Learning Foreign Languages : A History of Language Education, Assessment and Policy in Britain*, Routledge. <https://doi.org/10.4324/9781315624853>
- Mitchell, R. and Myles, F. (2019) 'Learning French in the UK Setting', *Apples – Journal of Applied Language Studies*, 13 (1), pp. 69–93. <https://doi.org/10.17011/apples/urn.201903011690>
- Opdenakker, M-C and Damme, J. (2000) 'Effects of Schools, Teaching Staff and Classes on Achievement and Well-Being in Secondary Education: Similarities and Differences Between School Outcomes.' *School Effectiveness and School Improvement*, 11 (June). [https://doi.org/10.1076/0924-3453\(200006\)11:2;1-Q;FT165](https://doi.org/10.1076/0924-3453(200006)11:2;1-Q;FT165)
- Sirin, S. R. (2005) 'Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research', *Review of Educational Research*, 75 (3), pp. 417–53. <https://doi.org/10.3102/00346543075003417>
- Slavin, R. E. (2010) 'Class Size and Student Achievement: Small Effects of Small Classes.' *Educational Psychologist*, June. [https://doi.org/10.1207/s15326985ep2401\\_4](https://doi.org/10.1207/s15326985ep2401_4)
- Stables, A. (1990) 'Differences Between Pupils from Mixed and Single-Sex Schools in Their Enjoyment of School Subjects and in Their Attitudes to Science and to School', *Educational Review*, 42 (3), pp. 221–30.  
<https://doi.org/10.1080/0013191900420301>
- Sullivan, A. (2009) 'Academic Self-Concept, Gender and Single-Sex Schooling.' *British Educational Research Journal*, 35 (2), pp. 259–88. <https://doi.org/10.1080/01411920802042960>

- Theobald, E. (2018) 'Students Are Rarely Independent: When, Why, and How to Use Random Effects in Discipline-Based Education Research', *CBE – Life Sciences Education*, 17 (3). <https://doi.org/10.1187/cbe.17-12-0280>
- Thomas, G. and Vaughan, M. (2004) *Inclusive Education: Readings and Reflections*. Inclusive Education. ERIC.
- Webster, R. and Blatchford, P. (2015) 'Worlds Apart? The Nature and Quality of the Educational Experiences of Pupils with a Statement for Special Educational Needs in Mainstream Primary Schools', *British Educational Research Journal*, 41 (2), pp. 324–42. <https://doi.org/10.1002/berj.3144>
- Wingate, U. (2018) 'Lots of Games and Little Challenge: A Snapshot of Modern Foreign Language Teaching in English Secondary Schools', *Language Learning Journal*, 46 (4), pp. 442–55. <https://doi.org/10.1080/09571736.2016.1161061>

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit <https://nationalarchives.gov.uk/doc/open-government-licence/version/3> or email: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk)


Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

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



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

<https://educationendowmentfoundation.org.uk>

 @EducEndowFoundn

 Facebook.com/EducEndowFoundn