



Education  
Endowment  
Foundation

# Best Practice in Grouping Students Intervention A: Best Practice in Setting Evaluation report and executive summary September 2018

## Independent evaluators:

Palak Roy, Ben Styles, Matt Walker, Jo Morrison, Julie Nelson and Kelly Kettlewell



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:**

**Danielle Mason**  
Head of Research

Education Endowment Foundation  
9th Floor, Millbank Tower  
21–24 Millbank  
SW1P 4QP  
p: 020 7802 1679  
e: [danielle.mason@eefoundation.org.uk](mailto:danielle.mason@eefoundation.org.uk)  
w: [www.educationendowmentfoundation.org.uk](http://www.educationendowmentfoundation.org.uk)

## About the evaluator

The project was independently evaluated by a team from the National Foundation for Educational Research (NFER). The principal investigator for this trial was Dr Ben Styles, Head of NFER's Education Trials Unit and the day-to-day trial manager was Palak Roy, Senior Trials Manager from NFER's Centre for Statistics.

### Contact details:

#### National Foundation for Educational Research (NFER)

The Mere  
Upton Park  
Slough  
Berkshire  
SL1 2DQ

**p:** 01753 637385

**e:** [p.roy@nfer.ac.uk](mailto:p.roy@nfer.ac.uk)

## Acknowledgements

We would like to thank the delivery team at UCL Institute of Education for their support throughout the evaluation. We are grateful to Sarah Lynch and Dr Anneka Dawson (in her previous role at NFER) for their help in undertaking some of the interviews in the first year of the evaluation and Sally Bradshaw for her help during the randomisation. Thanks are also due to Asma Ullah, Keren Beddow, Kathryn Hurd and Shalini Sharma in NFER's researcher operations for their efforts to maintain ongoing relationships with the schools. We are also grateful to Julie Thompson, Sagina Khan and Margaret Parfitt for their administrative support including formatting the report. Finally, we appreciate the support of the team at the Education Endowment Foundation, Emily Yeomans, Camilla Nevill, Calum Davey and from 2016 onwards, Dr Anneka Dawson for their guidance throughout this evaluation.

## Contents

Contents.....	2
Executive summary.....	3
Introduction .....	6
Methods .....	12
Impact evaluation .....	27
Process evaluation.....	42
Conclusion.....	52
References .....	55
Appendix A: EEF cost rating .....	58
Appendix B: Security classification of trial findings.....	59
Appendix C: Recruitment material.....	61
Appendix D: Randomisation syntax .....	79
Appendix E: Theory of Change .....	83
Appendix F: Fidelity measures .....	84

## Executive summary

### The project

The Best Practice in Setting intervention was designed by academics at UCL Institute of Education to improve the educational attainment and self-confidence of students in Years 7 and 8 who are currently placed in attainment groups for maths and/or English.

Setting (which we use to mean grouping students in classes by their current attainment levels) is practiced in England to a varying degree and the extent of the practice can be dependent upon subject, year group, or resource availability (Kutnick *et al.*, 2005). However, by Key Stage 3, setting is commonplace in mathematics (Ofsted, 2013) and is also increasingly introduced in English in the later stages of Key Stage 3 and as students enter Key Stage 4 (Ofsted, 2013).

Setting or streaming appears to benefit higher attaining pupils whilst being detrimental to the learning of lower attaining learners (see the Sutton Trust-EEF Teaching and Learning Toolkit; Slavin, 1990; and Sukhnandan and Lee, 1998). There are a number of practices associated with setting that may be the cause of this differential effect. This intervention was designed to raise pupil attainment by helping schools address these previously-identified poor setting practices through the application of *best practice principles*, including:

- assigning teachers randomly to sets to ensure all pupils received high quality teaching;
- focusing on pedagogy (through training and materials);
- having no more than four different sets for any subject, which resulted in broader sets and aimed to reduce hierarchies and any negative impacts of misallocation;
- assigning students to sets based on independent measures of prior attainment, rather than subjective teacher assessment; and
- providing the opportunity to reassign students to different sets as suggested by their attainment results over the year.

Teachers were trained in how to use the principles in their schools and then expected to apply them. The training was delivered by UCL Institute of Education researchers over the course of two academic years (2015/2016 and 2016/2017) and performance in English and maths tests was measured for a cohort of students that progressed from Year 7 to Year 8 during this time. The evaluation was set up as a school-randomised trial. One hundred and twenty seven schools were randomised to either receive the Best Practice in Setting intervention (in English and/or maths) or continue with their existing setting practices.

For the process evaluation, surveys and interview data were used to measure implementation and fidelity, explore organisational and attitudinal changes, and assess scalability.

### Key conclusions

1. The project found no evidence that the Best Practice in Setting intervention improves maths or English attainment for children in Years 7 and 8. For English, pupils in the intervention schools made slightly less progress than the control pupils, but this finding has no meaning given its level of uncertainty and very low security rating.
2. There was no evidence of impact on pupils' self-confidence in maths. For English there was a small positive difference in pupil self-confidence, but this result has no meaning due to its level of uncertainty and the large amount of measurement attrition.
3. School and teacher buy-in was low. Half of the schools in the maths trial and more than half of the schools in the English trial ceased intervention delivery before follow-up. Attendance at training sessions decreased over time (to 21% for the final maths session and 12% for the final English session). With this level of treatment attrition, being able to engage schools with the programme was demonstrated to be at least as important as the programme itself. Future interventions aimed at changing setting practices should be designed with the issue of engagement in mind.
4. The process evaluation revealed mixed views from participants. While some were largely positive about the intervention, many interviewees thought that what they were being asked to do represented little change from what they already do. Some schools reported that the intervention was onerous and difficult to deliver.
5. Some schools struggled to implement the intervention in full (for example, there was poor fidelity to teachers being randomly allocated to sets), this seemed to be due to schools finding it difficult to make some changes and therefore adapting principles to make them more implementable.

### EEF security rating

The trial was a large efficacy trial designed to test whether the Best Practice in Setting intervention can work under developer-led conditions in a number of schools. The trial was a well-designed randomised controlled trial. However, the findings for maths have a low to moderate security rating and the findings for English have a very low security rating. This is because 34% of the pupils who started the maths trial and 60% of the pupils who started the English trial were not included in the final analysis because their school did not provide test data. The pupils in intervention schools were similar to those in the control schools in terms of prior attainment and the null effects were largely robust to analyses that took into account the missing data.

### Additional findings

Due to the problems with intervention implementation by schools and the high attrition rates for follow-up testing, it is difficult to conclude anything certain about the impact of Best Practice in Setting from the results of this trial. There was no evidence that it improves attainment or self-confidence. The primary outcomes were attainment in English and maths and the secondary outcomes were self-confidence in English and maths; all measured in Year 8. Subgroup analysis investigated whether effects varied with free school meal eligibility (now or in the past) and prior attainment.

There are several key pieces of evidence that suggest that many schools did not 'buy in' to this intervention to the extent that they needed to in order for any impact to be seen:



- 30 (out of 61) maths intervention schools and 27 (out of 43) English intervention schools dropped out of delivering the programme. Attendance at workshops reduced dramatically during the intervention.

- Most interviewees thought that what they were being asked to do was no different from what they do normally. This view was, however, contrasted with poor fidelity to some practices, with data suggesting that some elements of the programme (for example random allocation of teachers to sets) were not implemented by the majority of schools. This seems to be due to schools adapting principles to make them more implementable. In its current form the programme does not result in schools making the prescribed adaptations to setting practices. Due to the low fidelity programme implementation by schools we cannot conclude that the underlying idea – optimising practices associated with setting to ameliorate poorer outcomes for those in lower sets – does not work, but this research demonstrates the challenges that schools face in adapting setting practices and the need to understand how best to support schools in this. The findings of this evaluation indicate that more attention needs to be paid to school requirements to inform future research into this aspect of common school practice.

## Cost

The average cost of the Best Practice in Setting intervention per department was around £2,500, or £14.04 per pupil per year when averaged over three years. The main costs were the cost of teacher training and programme materials. Schools also needed to meet the cost of staff cover for two days of teacher training per department for the first two years of the intervention.

**Table 1: Summary of impact on primary outcomes**

Group	Effect size (95% confidence interval)	Estimated months' progress	No. of pupils	P-value	EEF security rating	EEF cost rating
<b>Maths</b>	-0.01 (-0.11, 0.10)	0	2383	0.91		£££££
<b>English</b>	-0.08 (-0.21, 0.05)	-1	939	0.23		£££££
<b>Maths (everFSM only)</b>	0.02 (-0.14, 0.18)	0	713	0.80	N/A	£££££
<b>English (everFSM only)</b>	-0.01 (-0.25, 0.22)	0	265	0.90	N/A	£££££

## Introduction

This report presents findings from an efficacy trial and process evaluation of the Best Practice in Setting intervention, an intervention which trained schools in a best practice approach to grouping Year 7 and 8 students in English and mathematics by attainment. The intervention was developed by King's College London (KCL) and later transferred to University College London, Institute of Education (UCL Institute of Education).<sup>1</sup> The evaluation was conducted by the National Foundation for Educational Research (NFER) between May 2014 and November 2017.

## Background evidence

There are various forms of ability grouping, which potentially have different ramifications for students. These include streaming, setting by subject, and within-class ability grouping. These approaches are described by Taylor *et al.* (2017) as follows:

*Streaming refers to the practice of separating students according to a general notion of 'ability' across all (or a majority of) subjects so that students are taught in the same streamed groups for most lessons. Setting is more flexible and students are grouped according to 'ability' in individual subjects. In England, within-class 'ability' grouping is most commonly practiced in primary schools, where children are seated at 'ability tables' within a class containing a broad range of prior attainment.*

The full extent of streaming and setting across secondary schools in England is unclear, although in practice, schools use a variety of approaches, dependent upon subject, year group or resource availability (Kutnick *et al.*, 2005). In around one third of schools visited by Ofsted in 2013, students were taught predominantly in mixed attainment groups through Key Stage 3, with setting by ability generally only commonplace in mathematics (Ofsted, 2013). Setting was introduced at various points for English and science, but often only in the later stages of Key Stage 3. In response to a Parliamentary question in September 2016 asking whether the Department for Education held data on the use and prevalence of setting and streaming in schools in England, Nick Gibb, Minister of State for Schools, responded that: 'In-school organisation, such as streaming or setting, is a matter for individual schools, and data is not collected or held by the Department on this matter.'

Setting or streaming appears to benefit higher attaining pupils whilst being detrimental to the learning of lower attaining learners (see the Sutton Trust-EEF Teaching and Learning Toolkit; Slavin, 1990; and Sukhnandan and Lee, 1998). Or, as Slavin (1990) summarises, high achievers gain from ability grouping at the expense of low achievers. An additional perspective was provided by Andreas Schleicher in response to a question posed during a House of Commons Education Committee on Academies and Free Schools in 2015. He commented that, according to PISA data, when schools use setting for individual subjects selectively, and allow for movement between sets, this has no detrimental effect on attainment, but when they apply setting across all subjects—that is, implement streaming—this opens up socio-economic disparities.

Drawing on this evidence base, academics at UCL Institute of Education developed an intervention to help schools apply *best practice principles* when grouping students by attainment in order to overcome a variety of recognised challenges associated with streaming and setting. These challenges are described by the development team in the evaluation protocol (EEF, 2015) and include:

---

<sup>1</sup> The delivery team moved from KCL to UCL Institute of Education. Throughout the report, they will be referred as either the delivery team or UCL Institute of Education.



- **Inappropriate matching of teacher quality to student needs.** There is a tendency for teachers perceived as 'lower quality' to be placed with lower ability groups (Slavin, 1990).
- **Misallocation.** Certain groups of children—those from lower socio-economic groups, some minority ethnic groups, or with special educational needs (SEN)—are over-represented in low sets and streams (Kutnick *et al.*, 2005), and such allocation does not always match 'ability' as designated by test scores (Dunne *et al.*, 2007). Teachers and schools tend to underestimate the extent of student misallocation (Hallam and Ireson, 2005).
- **Lack of fluidity, or movement between groups.** Once placed in an ability group, students tend to remain there, irrespective of their progress or attainment (Flores, 1999; Dunne *et al.*, 2007). Teachers and schools tend to overestimate the extent of movement between groups (Hallam and Ireson, 2005).
- **Variable teacher expectations.** Teachers of high sets typically convey high expectations through provision of fast-paced and challenging work, whereas students in low sets often receive slow-paced teaching that covers less of the curriculum (Ireson *et al.*, 2005). Ability grouping can be confounded by teacher perceptions and assumptions about students' capabilities and attributes (including attainment levels, social class, gender and ethnicity).
- **Fixed pedagogy.** Teachers generally adopt limited teaching styles when teaching students in sets (Boaler *et al.*, 2000; Hallam and Ireson, 2005).
- **Impoverished curriculum and qualifications:** Ability grouping can produce an 'artificial ceiling', wherein students in lower sets are excluded from higher-tier study and qualification routes (Dunne *et al.*, 2007; Ireson *et al.*, 2005).
- **Polarised student engagement and attitudes:** The impact of ability grouping on student identities and expectations is thought to impact on outcomes (EEF Toolkit).<sup>2</sup> Ball (1981) observed that grouping created a polarisation of students into pro- and anti-school factions: those in low attainment groups became 'anti-school', with consequent detrimental impact on their achievement and aspirations. Recent research by Hallam and Ireson (2007) hints at continuing dissatisfaction among students in low sets, finding that nearly two-thirds (62%) of young people in bottom sets wished to change set.

The combination of these factors is argued to create a 'self-fulfilling prophecy' in which lower-set students gain restricted access to the curriculum, which in turn creates barriers to progressing through sets and to higher-tier study. Students' growing self-perceptions about their 'abilities' impact on their aspirations, effort and engagement levels.

The Best Practice in Setting intervention was designed by academics at UCL Institute of Education to improve the educational attainment and self-confidence of all students by ensuring that their progress benefitted from evidence-based good practice concerning setting. The intervention was designed to help schools address poor practices associated with setting through the application of *best practice principles* including:

- **random teacher assignment to sets**, to ensure that teacher quality did not relate to set level;
- **allocation of students to no more than four sets via independent measures** (attainment scores), to ameliorate hierarchies and the effects of misallocation, as well as to ensure greater opportunity for peer learning;

---

<sup>2</sup> <https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit>

- **review progress and reassign students to different sets (strictly on the basis of attainment results)** three times across the two-year period to ensure that any student performing above or below the set level are moved, and that consequently sets are rigorously based on attainment rather than other factors; and
- **a focus (through training, development and materials) on pedagogy**, challenging teachers' beliefs regarding students' 'fixed abilities', encouraging high expectations, ensuring equality of access to a rich and demanding curriculum, and facilitating schools to implement good practice in setting.

Full details of the intervention are provided below using the TidIER framework as adapted by the EEF (Humphrey *et al.*, n.d.c).

## The intervention

### Name:

Best Practice in Setting (BPS)

### Rationale:

Previous research has shown that students from disadvantaged backgrounds tend to be over-represented in lower attaining sets and streams (Dunne *et al.*, 2007). Students in these groups typically make less progress than their peers in higher attainment groups. In addition, lower sets and streams can be subject to elements of poor practice such as being taught a different curriculum to higher attaining students, being taught at a slower pace, receiving poorer quality teaching, and experiencing low expectations.

### Recipients:

The evaluation focused on teaching within English and maths with a cohort of Year 7 pupils who were then followed into Year 8. Of the 127 participating schools, 121 schools took part in the maths trial (61 intervention schools and 60 control schools) and 79 schools took part in the English trial (43 intervention schools and 36 control schools). Figure 3 presents details on the number of schools randomised to either conditions for single subject or both the subjects.

### Professional development and materials:

Schools were expected to support staff to attend four workshops run by the developer, and to develop, implement and monitor a school-wide strategy aimed at providing high expectations and a stimulating, challenging curriculum to all students.

- In the first workshop, the developers presented workshop participants with the research evidence behind the intervention together with instructions on set allocation and teacher attitudes and practices together with examples of how the schools in the developmental phase<sup>3</sup> addressed this. Activities supported reflections on the anticipated challenges and the opportunities presented by the intervention.
- The second workshop centred on the topics of high expectations, praise, and feedback. Information and materials drew on Dweck's work on 'mindset' (Dweck, 2008, 2010) and Ames's work on task-orientated rather than ego-oriented goals (Ames, 1992), along with other research showing that 'ability' should not be considered pre-determined or fixed, and the impact of teacher expectations on student outcomes.

---

<sup>3</sup> Note that the development phase was not part of this independent evaluation by NFER.

- The third workshop focused on the importance of rich curriculum to support student movement between sets. To help participants make the common curriculum accessible to all students, research evidence on best practice principles for differentiation was also presented and discussed.
- The fourth workshop was designed to reiterate the key messages ahead of the second year of the intervention, share achievements, problem-solve challenges, and plan for the year ahead.

**Procedures:**

In addition to the professional development on offer, schools were expected to follow a number of procedures in the way they allocated students to sets. Where possible, schools were asked to submit to independent randomisation of teachers across sets. Where schools were not prepared to adopt this approach, schools were asked to distribute teachers according to principles provided through the intervention (for example, placing some of their most experienced teachers with lower attaining sets and not assigning NQTs to low sets). Students were subject to independent measures of attainment, and set at the beginning of Year 7 accordingly. Schools were asked to adopt a maximum of four sets, and student progress and attainment was to be reviewed three times across the two-year period and students moved between sets accordingly. Headteacher support was expected for all of the above procedures. The process evaluation found that while most schools adhered to the set restrictions, few appeared to submit to the independent randomisation of teachers across sets, and some students were allocated to sets based on factors other than just their Key Stage 2 attainment data. These issues are discussed in the results section.

**Implementers and mode of delivery:**

The intervention was designed to be implemented by the teachers in participating schools, with delivery undertaken as part of the normal timetable. Four regional workshops were provided across the two-year intervention, with the expectation that approximately four<sup>4</sup> staff from each participating school (two teachers from each department, maths and English) would attend each workshop. (The delivery team identified a central location (regional hub) accessible to school staff after the school day.) The four workshops were delivered in July 2015, September 2015, February 2016 and October 2017 via twilight sessions delivered at the regional hubs. It was expected that one of these would be the relevant subject(s) Head of Department who would become the agent responsible for implementation, monitoring and working with teachers in their schools. Additional, 'bespoke' support and advice was provided to schools by the development team as necessary. While the process evaluation found that most participating English and mathematics departments were represented at each training workshop, overall, fewer teachers attended the workshops than was expected. This is discussed in the results section.

**Tailoring:**

The Best Practice in Setting intervention was a prescribed intervention, and optimal treatment fidelity was emphasised. Nonetheless, participating schools were permitted to make limited adaptations (for example, to the process of allocating teachers to sets) in order to accommodate local needs.

**Planning:**

---

<sup>4</sup> The protocol mentions that approximately six to ten staff from each intervention school were expected to attend the workshops, in contrast to the Memorandum of Understanding (MoU) (that was sent to schools) which mentions four staff to attend these workshops. Throughout the report, the findings are reported that are in line with the MoU.

Strategies to maximise implementation effectiveness included the use of initial developer-led workshops. The initial workshops aimed to support the implementation of department-wide approaches to addressing factors such as raising teacher expectations, providing a rich curriculum, and supporting differentiation of work. Later workshops were designed to review earlier learning and to create an arena in which participants could discuss issues arising, monitor progress, and share best practice. The process evaluation revealed that some participants found the workshops repetitive, while some would have liked the workshops to have focused less on theory and more on providing practical guidance on teaching through attainment setting. This is discussed in the results section.

## Evaluation objectives

The evaluation was set up as an efficacy trial with a target sample size of 120 secondary schools. Eligible pupils were in Year 7 at the start of the trial in the autumn term of academic year 2015/2016. The aim of this evaluation was to measure the impact of the Best Practice in Setting intervention on pupils' mathematics and English outcomes compared to the pupils from the control schools. This overarching aim was achieved by answering a number of research questions.

The primary research questions were:

1. What is the impact of Best Practice in Setting on pupils' attainment in mathematics?
2. What is the impact of Best Practice in Setting on pupils' attainment in English?

The secondary research questions were:

1. What is the impact of Best Practice in Setting on pupils' self-confidence in mathematics?
2. What is the impact of Best Practice in Setting on pupils' self-confidence in English?

Subgroup analysis on the primary outcome measures explored the differential impact of the intervention on pupils' attainment outcomes based on different prior attainment and whether or not the pupil was ever eligible for free school meals (FSM) in the previous six years.

The objective of the process evaluation was to examine implementation and fidelity to the intervention. It also aimed to identify factors that affect the impact of the interventions and might explain the findings of the quantitative evaluation. We aimed to look for evidence of effects and issues which would need to be considered for a wider roll-out.

## Ethical review

KCL obtained ethical approval for the study on 24 July 2014 and NFER's Code of Practice Committee approved the data collection on 5 March 2015. For the data collection and processing, this trial followed the EEF's advice on consent and the Data Protection Act (EEF, n.d.a). Schools opted into the trial through the headteacher (or their designated deputy) signing a Memorandum of Understanding (MoU) during recruitment. The school information sheet, invitation letter and reply form all contained relevant information about consent and how the NPD data would be used. Administrative pupil data (pupil names, dates of birth and UPN) was obtained from the participating schools in order to match assessment data and attitudinal data to the NPD background characteristics (prior attainment at Key Stage 2 and FSM eligibility). Parental opt-out consent letters were administered by schools prior to schools sending this data to NFER.

Appendix C provides the Frequently Asked Questions about BPS, NFER's recruitment invitation letter to schools, parent consent letter, school MoU and reply form.

## **Project team**

The principal investigator for this trial was Dr Ben Styles, Head of NFER's Education Trials Unit. The day-to-day trial manager was Palak Roy, Senior Trials Manager from NFER's Centre for Statistics (who took on this role from March 2015). Prior to this the trial was managed by Matt Walker. They were supported by NFER trial statisticians Sally Bradshaw and Jo Morrison. The process evaluation was led by a team of researchers from NFER's Centre for Policy and Practice Research: Matt Walker, Dr Julie Nelson and Kelly Kettlewell. The school recruitment and communications were managed by researchers from NFER's Research Operations department: Keren Beddow, Asma Ullah and Kathryn Hurd. The GL Assessment test administration was managed by Shalini Sharma and the tests were administered by trained NFER test administrators.

The intervention was developed and delivered by the team at KCL led by Professor Becky Francis who was at KCL until June 2016. After this point, the delivery team moved to UCL Institute of Education. The project manager was Dr Becky Taylor. They were supported by a team of researchers from Queen's University Belfast (QUB) and the University of Nottingham (UoN).

NFER was responsible for the trial design, school recruitment and ongoing relationship with the schools (jointly with the delivery team), randomisation, administrative pupil data collection, analysis and reporting of the independent evaluation.

The project was supported and guided by EEF staff Emily Yeomans, Camilla Nevill, Calum Davey and, from 2016 onwards, Dr Anneka Dawson.

## **Trial registration**

The trial was registered as trial number: ISRCTN17963123.

## Methods

### Trial design

School-level randomisation was necessary as ‘best practice’ encapsulates a whole-school approach, for example allocation of teachers across sets. The trial was run as a cluster RCT where schools were randomised to receive either the intervention or to be part of a control group. A total of 127 secondary schools were randomised and the trial followed children who were in Year 7 at the beginning of the trial until the end of their Year 8 (academic years 2015/2016 to 2016/2017). Schools allocated to the intervention group received the Best Practice in Setting intervention and schools allocated to the control group continued with their student grouping practices as usual. In order to keep the control group engaged with the trial, schools from this group received £1000 once they completed the primary outcome tests in summer 2017.

### Selection criteria

At recruitment, it was assumed—from researchers’ own experience, on the basis of the literature, and from the pilot study—that most schools were employing setting for both Key Stage 3 English and mathematics. However, to support recruitment, any school using ability grouping in Year 7 was eligible to take part, including schools that were undertaking streaming. Schools that stream (that is, allocate their pupils to fixed attainment groups across subjects) were eligible if they were prepared to amend their grouping arrangements to setting. The following eligibility rules were used to recruit schools and therefore no schools in the trial population employed mixed attainment grouping for the subject(s) in question.

**Table 2: Eligibility criteria to recruit secondary schools for the BPS trial**

Year 7		Year 8		Eligible?
English	Mathematics	English	Mathematics	
Setting	Setting	Setting	Setting	Yes
Mixed	Mixed	Setting	Setting	No
Mixed	Setting	Mixed	Setting	Yes (for mathematics outcome only)
Mixed	Mixed	Mixed	Setting	No
Mixed	Mixed	Mixed	Mixed	No
Streaming	Streaming	Streaming	Streaming	Yes, if prepared to set

Table 2 sets out scenarios where schools set in maths but not English; it does not include the reverse. This omission was deliberate and reflected researchers’ own experiences of working with schools that setting is more commonly practiced in maths than English. It was therefore anticipated that not all schools would be willing or able to participate in both English and mathematics. It was difficult to recruit

schools who were willing to set in English in Years 7 and 8. Furthermore, it was considered unethical to persuade schools to set if they employed mixed attainment teaching, given the evidence suggesting that mixed attainment teaching is beneficial to disadvantaged students. Of the 127 schools that took part in the trial, 121 took part in the maths trial compared to only 79 schools that took part in the English trial (see randomisation section for further details). Of these 79 schools for the English trial, six schools took part only in English. Separate sample sizes and power calculations were considered for both the subjects and outcome measures.

## Recruitment

The recruitment for this trial was jointly managed by NFER and KCL, each deploying their own distinct recruitment approach. KCL publicised both the trials in social and traditional media. Schools then contacted them to inquire more about this trial as well as the pilot trial on Best Practice in Mixed Attainment (BPMA). One hundred and seventy two schools made contact with KCL.

In contrast, NFER contacted schools by drawing a stratified random sample of schools to take part in one or the other trial. For this trial, NFER's sample included all English state secondary schools that had both year groups present: Years 7 and 8. Moreover, schools with fewer than 60 pupils were excluded from the sample (as this was the required number of pupils to be tested from each school; please see sample size section for more details). As both the organisations were recruiting concurrently, NFER excluded schools from 44 Local Authorities (LAs) as KCL focused their recruitment in these areas. These LAs were spread out across England. On this basis, NFER drew a sample of 639 state secondary schools that was stratified by government office region, school type and proportion of pupils eligible for free school meals (FSM). We approached all schools with an initial invitation letter to the headteacher in March 2015. This was followed by an extensive reminder exercise which consisted of a mixture of written reminder letters, phone reminders and emails to named individuals where available. In most cases, NFER staff asked to speak to the headteacher or a senior leader. After a month, we amended the approach slightly, contacting subject heads directly in the first instance, as it was proving to be difficult to establish contact with headteachers. We therefore sent recruitment letters to the Head of Maths and/or Head of English, who themselves sought headteacher buy-in. We also asked to speak to subject heads when phoning the school.

By June 2015 we decided to draw another sample as the recruitment target was not achieved. NFER drew another stratified random sample of state secondary schools with similar inclusion/exclusion criteria as the main sample. The main focus of this recruitment effort was to achieve the total recruitment target, but also to facilitate the feasibility of intervention delivery. The delivery team identified five possible regional hubs where the professional development sessions were going to take place and LAs were selected based on geographical proximity. This resulted in NFER approaching an additional 195 schools from 13 LAs.

The final list of participating schools for the trial was the total number of schools recruited by NFER and KCL. These schools were then put forward for randomisation.

## Outcomes measures

### Primary outcome measures

Some schools took part in the trial for both subjects whilst other schools took part only for one subject. This means there were slightly different schools taking part in maths to those taking part in English. Hence, it was necessary to undertake two separate analyses, one for each subject. It was further decided during the Statistical Analysis Plan (SAP) (EEF, 2017) review stage that the pupils taking English and maths tests are different and therefore adjustment for multiple testing is not required in this case.

As the intervention was aimed at pupils in Years 7 and 8, testing was necessary as there is no statutory assessment in these years. Testing took place in Year 8 after two years of the intervention. GL Assessment's Progress Test in English<sup>5</sup> (PTE13) and Progress Test in Mathematics (PTM13) were used. These primary outcome measures of attainment were used to answer the following research questions:

1. What is the impact of best practice in setting on pupils' attainment in mathematics?
2. What is the impact of best practice in setting on pupils' attainment in English?

As these tests have a broad coverage of the curriculum, the raw total score for each subject that covers all curriculum content was used. Maths total score (maximum possible score 70) consisted of fluency in facts and procedures, fluency in conceptual understanding, mathematical reasoning and problem solving. English total score (maximum possible score 66) consisted of spelling, grammar and punctuation, reading comprehension: narrative and non-narrative. For both these outcomes, higher scores indicate better performance. Further information for both the tests can be found on the GL Assessment website.<sup>6</sup>

NFER took responsibility for collecting and delivering the PTE13 and PTM13 in paper form using its test administrators. The test administrators had clear guidance from NFER on how to administer the test, which also emphasised the importance of the trial methodology. The administrators did not know the group schools belonged to and the administration guidance specifically asked them to avoid discussion with the school staff about the group allocation. Once the tests were completed, they were sent to GL Assessment for marking. Markers were also blind to group allocation.

Due to the lower than anticipated response rate for the GL Assessment test, NFER put in place extensive reminding strategies and also extended the time period for the test administration. A number of schools were also offered to test their whole Year 8 cohort as some schools responded that administering the test with 60 pupils was inconvenient as they could be in different classes. These schools were offered extra test scripts for free. Five schools were already planning to administer the GL Assessment tests with their entire Year 8 cohort and they preferred NFER to access their pupil results directly from GL.

There were 79 schools that took the GL Assessment tests; 76 in maths and 37 in English. In addition, there was one school that took the online versions of PTE and PTM tests. As the online assessments are not directly comparable to the paper versions of the assessment, it was necessary to link the raw scores for these assessments. As the curriculum domains covered by the paper and the online versions are the same and the age-standardised scores were standardised using national samples at the same time, they can be used to link the age-standardised scores from the online versions to the raw scores for the paper-based versions. Using pupils' age in months and their age-standardised scores on the

---

<sup>5</sup> At the time of the protocol, these tests were being developed by GL Assessment. They were being called New Progress in English (NPIE) and New Progress in Mathematics (NPIM). After the development, these tests were named Progress Test in English and Progress Test in Mathematics.

<sup>6</sup><https://www.gl-assessment.co.uk/media/1346/ptm-technical-information.pdf>

<https://www.gl-assessment.co.uk/media/1366/pte-technical-information.pdf>

<https://www.gl-assessment.co.uk/media/1384/ptm13-links-to-nc.pdf>

[https://www.gl-assessment.co.uk/media/1349/pte13-links-to-national-curricula\\_0.pdf](https://www.gl-assessment.co.uk/media/1349/pte13-links-to-national-curricula_0.pdf)



digital versions, we looked up the raw scores on the paper test versions that would have been obtained by a pupil of the same age with the same age-standardised score. In cases where the exact age-standardised score did not exist for that age,<sup>7</sup> linear interpolation was used. The primary analysis was done using only the paper tests. A subsequent analysis included the school that used the digital version, with its scores converted to 'paper' scores.

### **Secondary outcome measures**

As outlined in the protocol, the secondary research questions were:

- What is the impact of best practice in setting on pupils' self-confidence in mathematics?
- What is the impact of best practice in setting on pupils' self-confidence in English?

These were measured by administering a pupil survey at the start of Year 7 in September 2015 (baseline survey administered post randomisation)<sup>8</sup> and at the end of Year 8 in summer 2017 (follow-up survey). The delivery team was responsible for administration of these surveys. The surveys were administered with an entire cohort from all participating schools. However, for trial purposes, pupil survey data was analysed based on the original randomisation group and subject participation. For example, if a school was taking part in the mathematics only trial, pupils' self-confidence in English from this school was not considered in the English analysis.

In partnership with QUB, UCL Institute of Education developed pupil self-confidence measures in maths and English. Items were drawn from several instruments previously used: SDQII (Marsh, 1990); TIMSS questions (IEA, 2011) and PISA questions (OECD, 2012). The measures were developed using factor analysis on selected items from the baseline pupil survey data (combined dataset for both the trials, BPS and BPMA). The object of factor analysis is to reduce the number of variables required to explain the data from the original large number to a smaller set of underlying 'factors' which can be related to the original variables.

---

<sup>7</sup> For example, if scores were 98, 100, 101 and 104 for successive raw score points, it was not possible to look up values of 99, 102 and 103.

<sup>88</sup> Randomisation took place in summer 2015 and the first opportunity to administer pupil surveys with Year 7 pupils would be in the academic term autumn 2015 which was post- randomisation.

**Table 3: List of items included in the secondary outcome measures of self-confidence**

Composite measure	Constituent items	Source
Self-confidence in mathematics	Work in maths lessons is easy for me	Adapted from Marsh (1990) 'verbal self-concept'
	I am not very good at maths	Adapted from Marsh (1990) 'verbal self-concept'
	Maths is one of my best subjects	Adapted from Marsh (1990)
	I hate maths	Adapted from Marsh (1990)
	I do well at maths	Adapted from Marsh (1990) 'school self-concept'
	I get good marks in maths	Adapted from PISA self-concept in mathematics and Marsh (1990) 'verbal self-concept'
	I learn things quickly in maths lessons	Adapted from TIMSS self-confidence in learning mathematics and Marsh (1990) 'verbal self-concept'
Self-confidence in English	Work in English lessons is easy for me	Adapted from Marsh (1990) 'verbal self-concept'
	I am not very good at English	Adapted from Marsh (1990) verbal self-concept
	English is one of my best subjects	Marsh (1990) 'verbal self-concept'
	I hate English	Marsh (1990) 'verbal self-concept'
	I do well at English	Adapted from Marsh (1990) 'school self-concept'
	I get good marks in English	Adapted from Marsh (1990) 'verbal self-concept'
	I learn things quickly in English lessons	Adapted from Marsh (1990) 'verbal self-concept'

Table 3 lists items included in the factor analysis. All the items were retained and the composite measures were created as an average of all constituent items. Self-confidence in maths had a high internal reliability (Cronbach's  $\alpha$ ) of 0.88 and self-confidence in English had an internal reliability (Cronbach's  $\alpha$ ) of 0.86. Scores for these composite measures ranged from 1 to 5 with higher scores reflecting higher self-confidence in a given subject.

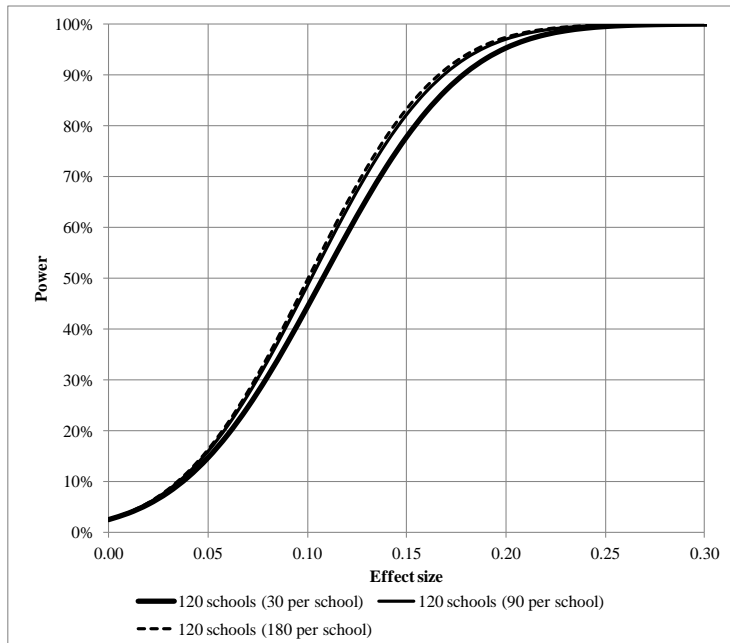
### Sample size

A number of within-school sample sizes were considered for analysis. In order to reduce the testing burden on schools without a sizeable impact on power, we decided that NFER would randomly select 60 pupils from the Year 8 school roll from each of the recruited schools. Half of the pupils would sit the mathematics test and half would sit the English test. For schools that were randomised for only one subject, 30 pupils would sit the test in that subject.<sup>9</sup> Some of these options are illustrated in the power curves in Figure 1. The MDES for the 30-students-per-school design used in the trial was 0.16. These power curves use the following assumptions: intra-cluster correlation of 0.15 (reduced from 0.2 through the use of Key Stage 2 as a covariate) and correlation between Key Stage 2 and Year 8 test of 0.7.

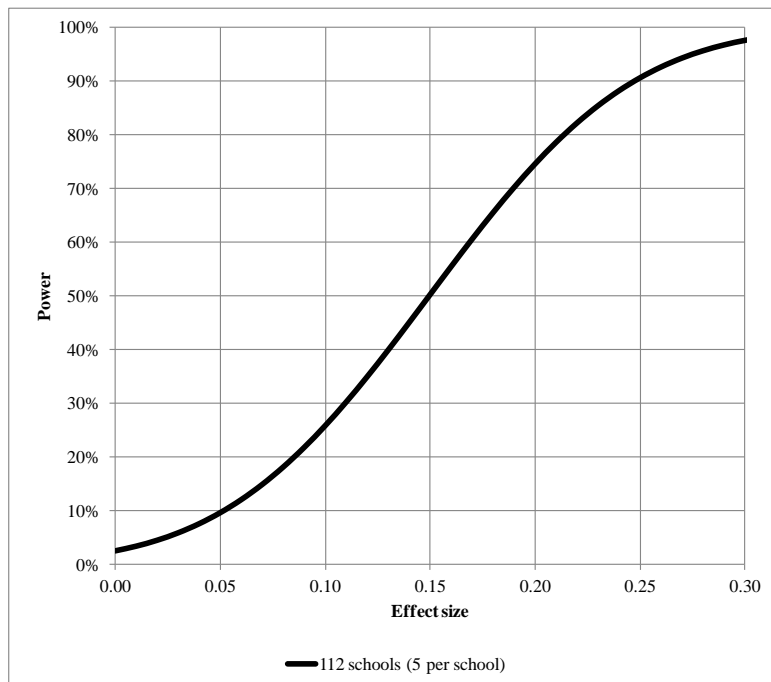
<sup>9</sup> Pupils being tested in English will be different from those being tested in maths. Sampled pupils will not be replaced in any case even if a pupil was no longer available for testing.

These assumptions were based on internal unpublished analyses by NFER statisticians carried out before NPD restrictions prevented such exploratory work.

**Figure 1: Power curves for 60 versus 60 schools (recruitment target as per the protocol)**



These power curves clearly illustrate how testing burden per school can be greatly reduced through within-school sampling with minimal impact on power. For such strategies to work, the within-school sampling has to be random to ensure unbiased cluster mean estimates. Within-school sampling has an impact on the power of subgroup analysis. As FSM-eligible pupils represent a particularly important subgroup, the power of a separate FSM analysis was considered at this stage. In order to achieve sufficient number of pupils with FSM eligibility, this within-school sampling was stratified by pupil FSM status, as obtained from NPD. Based on this calculation, we expected an average of five FSM pupils to be sampled in each school cohort and at least one FSM pupil to be sampled in each school in 93% of recruited schools. As we are just estimating regression coefficients, some small cluster sizes will not compromise the multi-level models (Snijders *et al.*, 2005). The minimum detectable effect size (MDES) was then calculated for FSM-only analysis at 0.22 for 112 schools (93% of 120) and with the same assumptions as above. This is illustrated in Figure 2 below.

**Figure 2: Power curves for FSM analysis (as per the protocol)**

## Randomisation

The intervention is a whole-school approach including, for example, allocation of teachers across sets, therefore school-level randomisation was conducted. Due to the difficulty in recruiting enough schools that wished to set in English, three different types of schools were identified: those taking part in both subjects, those taking part in maths only, and those taking part in English only. Therefore, it was necessary to stratify the randomisation by setting practice (English/maths/both) to allow a lower powered analysis to be run on English outcomes.

Randomisation was carried out by a statistician at NFER using a full SPSS syntax audit trail (see Appendix D) in June-July 2015 over a period of six weeks. The statistician was not blind to the randomisation groups. This was done in five blocks due to staggered school recruitment and intervention workshops running concurrently. At the time of writing the protocol, we envisaged carrying out the randomisation only once after the recruitment target was achieved. However, slower recruitment meant that fewer schools had signed up to the trial and were ready to be randomised in June 2015. At this time, first randomisation took place in order to ensure that there was sufficient time to organise the training workshops for these schools. Recruitment continued after this point and four more blocks of randomisation took place until the end of the summer term when the recruitment target was achieved. Details on the blocked randomisation are included in Table 4. After each block of randomisation, we informed the schools about their group allocation to aid intervention delivery.

There were 129 secondary schools randomised for this trial. One of these schools withdrew participation without the knowledge of group allocation and another school was randomised due to an administrative error. As these two schools withdrew prior to knowledge of group allocation, although randomised, they were considered as unbiased dropouts. Of the 127 remaining schools, 73 schools took part in both maths and English, 48 schools took part in maths only, and 6 schools took part in English only. This meant there were 121 schools taking part in the maths trial (61 intervention schools and 60 control schools). Of the 79 schools taking part in the English trial, 43 schools were assigned to the intervention group and 36 schools were assigned to the control group. Overall, there is an imbalance in the group allocation for schools taking part in the English trial. This occurred as a result of not correcting the group

imbalance that arose at each block. We deliberately adopted this approach to ensure that it was not possible to predict the sequence of group allocation.

**Table 4: Number and proportion of schools randomised**

		Intervention	Control
Block 1 (16 June 2015)	Maths n (%)	47 (49%)	49 (51%)
	English n (%)	30 (52%)	28 (48%)
Block 2 (6 July 2015)	Maths n (%)	11 (50%)	11 (50%)
	English n (%)	10 (56%)	8 (45%)
Block 3 (8 July 2015)	Maths n (%)	1 (100%)	0 (0%)
	English n (%)	1 (100%)	0 (0%)
Block 4 (14 July 2015)	Maths n (%)	1 (100%)	0 (0%)
	English n (%)	1 (100%)	0 (0%)
Block 5 (24 July 2015)	Maths n (%)	1 (100%)	0 (0%)
	English n (%)	1 (100%)	0 (0%)
<b>Total</b>	<b>Maths n (%)</b>	<b>61 (50%)</b>	<b>60 (50%)</b>
	<b>English n (%)</b>	<b>43 (54%)</b>	<b>36 (46%)</b>

## Analysis

We followed EEF analysis guidelines<sup>10</sup> and the published Statistical Analysis Plan (SAP) for the BPS trial. This section provides an overview of the analysis undertaken (the SAP provides further detail, see EEF, 2017).

### Primary intention-to-treat (ITT) analysis

The primary outcome analysis was ‘intention-to-treat’. There were two separate analyses, one for each subject. Pupils from schools that took part in a given subject trial (the maths trial or the English trial) were included in each analysis irrespective of compliance to the intervention. If a school took part in

<sup>10</sup>[https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing\\_a\\_Research\\_Report/2015\\_Analysis\\_f or\\_EEF\\_evaluations.pdf](https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_f or_EEF_evaluations.pdf)

both subjects, its pupil data was included in both the analyses, maths and English. Two separate multilevel models were run with two levels (school and pupil) so the analysis could account for cluster randomisation. The raw total score in mathematics for PTM13 was the dependent variable for the mathematics model. The model included, as covariates, an indicator of whether the pupil is in an intervention school, pupil prior attainment as measured by Key Stage 2 maths point score (KS2\_KS2MATPS variable on NPD), and an indicator of whether the school took part in one subject or both (representing the stratification variable used at randomisation). Similarly, the raw total score in English for PTE13 was the dependent variable in the English model. This model again included as covariates an indicator of whether the pupil is in an intervention school, pupil prior attainment as measured by Key Stage 2 English point score (KS2\_KS2READPS variable), and an indicator of whether the school took part in one subject or both (representing the stratification variable used at randomisation).

The primary ITT models were ‘completers’ analyses where a pupil record was not included in the model if they had one or more missing values for the variables in the model. This means that pupils with missing data, either on outcome measures or any of the covariates, were not included in the analysis. Note that we did not include a variable to indicate whether a school had been recruited by NFER or KCL. Although the different recruitment approaches used might have meant there was a difference in the propensity for schools to respond to the intervention, we did not feel it was critical for the analysis. Randomisation protects us from this kind of unknown bias.

The numerator for the effect size calculation was the coefficient for the intervention group as identified in each model. The effect size was calculated using the total variance from the model without covariates as the denominator—equivalent to Hedges’ *g*. Confidence intervals for the effect size were derived by multiplying the standard error of the intervention group model coefficient by 1.96 and converting this value to an effect size.

In addition to the above models, we also report a point estimate (without a confidence interval) from similar models which don’t include the stratification variable. This is reported for the purposes of cross-study comparisons.

### **Imbalance at baseline**

Due to high attrition (37% and 53% schools lost to follow-up in the maths and English trial respectively), it was important to examine any bias that might have resulted. We examined the imbalance in the final samples using school and pupil background characteristics. Pupil characteristics included pupil everFSM status and prior attainment at Key Stage 2. We used multilevel modelling to examine imbalance and reported differences in prior attainment as effect size.

### **Missing data**

We ran two multilevel logistic models (one for each subject) with two levels (school and pupil) on whether or not a pupil was missing at follow-up, regressed on a number of covariates in addition to the ones in the main model.

Since there were many schools that withdrew participation from the primary outcome measurement, it was important to explore the level of missing data and the extent of bias. Since the extent of school dropout was unequal between randomised groups, we needed to conduct sensitivity analyses. This was done by initially running multilevel<sup>11</sup> multiple imputation and then extending this model using a weighting approach according to Carpenter *et al.* (2007). This approach works by replacing a simple

---

<sup>11</sup> Please note that for the English outcome, it was not possible to run a multilevel model due to the high proportion of missing data (this is a deviation from the SAP).

average by a weighted average where estimates from the imputations that are more likely under 'missing not at random' are considered more important. After adjusting for the observed variables, the chance of observing the outcome measure per unit change in that measure has log-odds ratio of  $\delta$ . If data are 'missing at random',  $\delta$  will be zero. If  $\delta$  is positive, the chance of observing the outcome measure is higher for higher values of the outcome measure.

We ran the ITT models using each of the imputed datasets (these datasets included imputed values for missing covariates as well as the outcome measures), and the results were pooled to give coefficients and standard errors that took account of the variance during imputation. These were compared with the original ITT models. While exploring further the assumption that 'data is missing at random', we adjusted the values of the outcome measures in the imputed datasets in a sensitivity analysis by changing the values of  $\delta$ .

### **Secondary outcome analyses—self-confidence models**

Models for the secondary outcomes of self-confidence were run similar to the primary outcomes of attainment with two dependent variables in two separate multilevel models, each including pupils from schools that took part in the relevant subject trial. The covariates for these models were similar to the primary models wherein pupil self-confidence measures in a given subject at baseline were one of the covariates instead of prior attainment measures.

### **Compliance**

The developer collected data on the level of school engagement throughout the two-year delivery period using a number of predefined variables as described in Table 5. They sent us data on these individual variables and we summarised them according to pre-agreed categorisation. This categorisation yielded three measures listed in the table. Measures for English are provided below as an example. Similar information was collected and reported for maths.

As specified in the Statistical Analysis Plan, we planned to run Complier Average Causal Effect (CACE) analysis by using measure two from Table 5 (effectiveness of setting/allocation practices). Unfortunately, the composite measure didn't have sufficient data over time for all intervention schools. Similar measures from control schools were also not available and therefore we incorporated this fidelity data with our own qualitative data that was collected via interviews with school staff. A more detailed account of the intervention fidelity is included in the process evaluation section.

**Table 5: Fidelity measures for the English outcome**

Combined measure	Variable	Level of measurement
1. Effectiveness of training practices	1. English department represented at each training session	Binary. Did the expected number and type of staff attend each session?
		0 = No
		1 = Yes
	2. Training is cascaded to members of the English department	Binary. Has some form of cascading/internal training taken place?
		0 = No
		1 = Yes (if one or more departmental members concur)
2. Effectiveness of setting/allocation practices	3. Setting arrangements follow BP principles (only 3–4 sets)	Binary.
		0 = 5 or more sets
		1 = 3/4 sets or fewer
	4. Teachers are randomly allocated to classes	Were BP principles followed?
		0 = No, teachers not randomised or allocated with reference to BP principles
		1 = Partial, teachers allocated with reference to BP principles 2 = Yes, teachers randomised to classes
	5. Students are allocated to classes according to KS2 results	Binary – 95% or more of students are allocated on the basis of KS2 results
		0 = No
		1 = Yes
	6. Students are re-set no more than three times in two years.	Binary – 95% or more of students are re-set no more than three times in two years (i.e. 95% or more of students are re-set 'at the most' three times in two years. Schools that re-set students more than three times in two years will get a value of 0)
		0 = No
		1 = Yes
3. High expectations	7. Teachers have high expectations for all students	Binary
		0 = No
		1 = Yes

**Subgroup analyses**

Subgroup analyses on the primary outcomes were carried out. We explored the differential effect for different pupil ability levels and pupils' FSM eligibility. An interaction term was added to the main models. The intervention indicator was interacted with pupil ability as measured by prior attainment measures at Key Stage 2 and the EVERFSM\_6 variable in separate models.



A separate analysis of FSM-only pupils was also carried out as per the EEF analysis guidance. These models were similar to the main models of overall effect but only included pupils who were eligible for FSM as measured by the EVERFSM\_6 variable.

### Implementation and process evaluation

A light touch process evaluation<sup>12</sup> ran alongside the impact evaluation to collect data on how far the best practice approach was being adopted and implemented by trained schools, and to explore any organisational and attitudinal changes occurring for schools and teachers as a result of the training and best practice approach. The information collected supported consistency and fidelity checks and helped to determine the extent to which the intervention is scalable.

The process evaluation collected information on:

- the resources involved in the intervention, including the training sessions, teacher time and staffing;
- the starting points of schools in setting in English and mathematics at Key Stage 3 (including clarification of setting criteria, streaming, etc.);
- changes made to school organisation (e.g. staffing allocations, material resources), curriculum and pedagogy (e.g. level of demand/challenge), and grouping criteria (e.g. how schools addressed misallocation, fixed positioning and movement of students between sets); and
- brief perceptual feedback on changes in school ethos, and teachers' attitudes, expectations and assumptions around ability grouping.

The process evaluation involved four strands of activity spread across three academic years, as detailed below.

---

<sup>12</sup> Qualitative and quantitative data was also collected and analysed by the delivery team; this was published separately.

**Table 6: Overview of process evaluation research strands**

Activity	Year 1 (2014/2015) (Developmental phase)	Year 2 (2015/2016)	Year 3 (2016/2017)
<b>Exploratory scoping interviews with the developer (KCL/UCL Institute of Education)<sup>13</sup></b>	✓		
<b>Attendance at training/briefing events for schools</b>		✓	
<b>Collection of schools' cost data (via school proformas administered by UCL Institute of Education in 2015/2016 and by NFER in 2016/2017)</b>		✓	✓
<b>Collation and review of project documentation</b>	✓	✓	✓
<b>Telephone interviews with Heads of English and mathematics</b>	✓	✓	✓

In the first year of the evaluation, exploratory scoping interviews were undertaken with three members of staff from the UCL Institute of Education team to identify relevant management and monitoring information. A Theory of Change (ToC) was also developed by the UCL Institute of Education team, which is included in Appendix E.

Researchers from the process evaluation team attended a London-based training event delivered by the development team in July 2015, prior to the trial starting in the autumn term. This enabled the evaluation team to gain a better understanding of the intervention and teachers' responses to it.

At the start of the intervention, information was collected to establish schools' ability-grouping practices prior to randomisation. This information was collected via reply forms that schools sent back (along with a signed MoU) to the recruiting organisation. Following randomisation, a further round of data collection was undertaken to establish details of the nature of implementation practices including any organisational, curricular and grouping changes that had been made. Throughout the intervention, information was collected by UCL Institute of Education—using staff questionnaires and proformas—on: schools' setting arrangements; the extent to which teachers were randomly allocated to classes; the extent to which students were allocated to classes according to Key Stage 2 results;<sup>14</sup> the frequency with which students were 're-set'; and the extent to which teachers reported having high expectations for all students.

In the second and third years of the evaluation, telephone interviews were undertaken with a sample of Heads of English and/or mathematics, or a suitable alternative spokesperson, such as a school senior leader, in the intervention schools. The interviews explored schools' reasons for participating in the trial, the extent to which schools were implementing each of the key intervention principles, how staff had responded to the intervention, and their perceptions of its impact. The interviews were undertaken using

<sup>13</sup> These interviews were meant to take place in Spring 2015 only as per section 4.4.1 in the protocol. The tick marks against 4.4.1 in Table 2 of the protocol don't match with the text in section 4.4.1 and should have been removed when the protocol was amended.

<sup>14</sup> NFER collected data on pupil groups along with administrative pupil data. This was matched with KS2 data by UCL Institute of Education to establish whether allocation was according to KS2 results.

a semi-structured research instrument. The target was to achieve interviews with staff in one or both departments from ten randomly selected schools in each year of the evaluation.

- In the second year of the evaluation, a total of 11 interviews were undertaken between March and June 2016, covering 13 different departments (six mathematics and seven English) across eight schools (two senior leaders were able to talk about both English and mathematics as the intervention has been delivered in both subjects in their schools).
- In the third year of the evaluation, a total of five interviews were undertaken between March and May 2017, covering six different departments (two mathematics and four English) across four schools (one senior leader was able to talk about both English and mathematics as the intervention has been delivered in both subjects in their school).
- Three staff from three schools were interviewed in both the second and third year of the evaluation.

As the numbers suggest, it was much harder to secure interviews with participating teachers in the third, rather than the second, year of the evaluation. This was despite repeated email and telephone reminders to schools, drawing a top-up sample of schools, and the offer of reducing interview time. Brief conversations with staff from schools that withdrew from the evaluation, together with the interviews with participants who remained, provided a range of potential explanations for this lower-than-hoped-for response rate. These included: a perception, among at least some school respondents, that the intervention was not a big change from their usual practice (eight interviewees), or was, conversely, over-burdensome and that they did not wish to participate in the evaluation as a result (two respondents by phone); changes in school grouping practices—from setting to mixed attainment grouping in the third year of the evaluation (one interviewee); or reduced Year 7 intakes in the third year of the evaluation, making continued participation unfeasible (one respondent by phone). The views of teachers are explored in the process evaluation section of this report.

## Cost analysis

Cost calculations are based on information that was supplied by the delivery team as well as the schools themselves. The project delivery team provided information about staff time and other direct, marginal costs, such as staff travel costs and venue hire. Data on teacher travel costs was also provided. This information was supplied on a termly basis. All intervention schools were asked to provide cost data but only a small number of schools supplied this information. A sub-sample of 16 schools in year one of the intervention, and 20 schools in year two, provided information on the number of hours of cover that were used, the number of these hours that required paid supply cover, and the cost of this, together with details of other monetary costs such as travel and subsistence costs. These school costs were for activities delivered in support of the intervention, but they were not directly associated with attendance at the training delivered by UCL Institute of Education. This information was collected using proformas, first by the delivery team in the summer term 2016, and then by NFER in the summer term 2017.

We estimated a per school cost based on the average costs for one department (English or mathematics) per year. Where both the English and mathematics department was participating in a school, the aggregated cost data was halved to provide a mean average for each participating department. Once a mean cost per department per year had been calculated, this figure was then divided by the average number of Year 7 pupils in participating schools (using administrative student data collected by NFER) to calculate the average cost per pupil per year. Cumulative costs are calculated and reported over a period of three years as per EEF guidance. More detail is provided in the analysis section.

## Timeline

The developmental phase commenced in September 2014 where the intervention was piloted with three schools.<sup>15</sup> The intervention delivery for the main trial commenced in September 2015. The primary attainment outcomes were captured through testing in summer 2017 and the first draft report was submitted in November 2017. Table 7 summarises the timeline for evaluation activities.<sup>16</sup>

**Table 7: Timeline**

Date	Activity
September 2014–June 2015	Developmental phase—schools implemented Best Practice in Setting approach
Jan–March 2015	Write, agree and publish the protocol
April–May 2015	Develop ToC(s) with providers Amended version of the protocol published
May–August 2015	School recruitment for the main trial completed School randomisation Trial registration on <a href="http://www.controlled-trials.com/">http://www.controlled-trials.com/</a> Interviews with the delivery team Professional development (PD) sessions commenced (1 <sup>st</sup> session July 2015) Observation of the PD sessions
Autumn term 2015	2 <sup>nd</sup> PD sessions Administration pupil data collection commenced (along with the parental opt-out letters) Baseline pupil and teacher survey administered
March–June 2016	Telephone interviews with sub-sample of participants
January 2016	NPD data request sent
February 2016	3 <sup>rd</sup> PD sessions
September 2016	Cost proforma sent to schools (school costs associated with 1 <sup>st</sup> academic year)
October 2016	4 <sup>th</sup> PD sessions
March–May 2017	Telephone interviews with sub-sample of participants
June 2017	Cost proforma sent to schools (school costs associated with 2 <sup>nd</sup> academic year) End-point pupil survey administration Administrative pupil data collection completed
May–August 2017	Write and agree SAP
July–September 2017	GL Assessment test administration and marking
August–November 2017	Analysis and reporting

<sup>15</sup> This phase was not evaluated by NFER.

<sup>16</sup> The timeline excludes aspects of the data collection that were the sole responsibility of the delivery team.

## Impact evaluation

### Participants

As explained in the selection criteria, schools had to meet the eligibility criteria for each subject trial separately. A total of 129 schools were recruited and randomised to take part in maths and/or English BPS trials. Two schools withdrew participation<sup>17</sup> without the knowledge of group allocation. These schools are removed from the trial analysis as they can be considered as unbiased dropout. Of the 127 trial schools, 121 took part in the maths trial and 79 took part in the English trial. Randomisation was stratified by setting practice (English/maths/both) where 65 schools were allocated to the intervention group and 62 schools were allocated to the control group. Please see the 'Allocation' section in Figure 3 for details about the group allocation for each subject trial.

The delivery team kept a record of the number of schools dropping out from the intervention at various points throughout the trial. Of the 61 maths intervention schools, 30 schools withdrew from the intervention at some point during the two years, and of the 43 English intervention schools, 27 withdrew from the trial over the two year period. Of the total 65 intervention schools, 36 schools withdrew participation from at least one subject trial. However, some of these schools agreed to complete the primary outcomes tests. For further details, see the section 'Drop-out from intervention' in Figure 3.

Pupil administrative data was required to be collected from schools in order to match pupil records to the NPD and the primary outcome measures. As the cohort for the trial was Year 7 pupils at the beginning of the intervention (in the academic year 2015/2016), it was not possible to collect this administrative data from schools any time prior to September 2015 as schools would not know the pupil details. This meant that the pupil data collection took place after the randomisation and any dropout at this point can be considered a biased dropout. A total of 96 schools provided NFER with their pupil data (42 intervention schools and 54 control schools). The remaining 31 schools didn't provide the pupil list and withdrew completely from the trial. We requested the de-identified NPD data for Year 7 pupils at these schools and they were marked as lost to follow up. The total number of pupils in Year 7 at the start of the trial is given in the 'Allocation' section of Figure 3. Before supplying the administrative pupil data to NFER, schools removed pupils whose parents opted them out of the data collection. This way, we could ensure that these pupils were never included in any future data collection or data matching with the NPD.

As discussed earlier, NFER collected the primary outcomes data by administering GL Assessment tests when the trial pupils were in Year 8 (in the second year of the trial, academic year 2016/2017). For most schools that took part in the testing, we sampled 30 pupils randomly to take each subject test such that 60 pupils would sit the GL tests from a school taking part in both the subjects. This selection was stratified by pupil everFSM (as measured by EVERFSM\_ALL\_SPR15 variable from the NPD). A small number of schools preferred to test an entire Year 8 cohort. To summarise, the following types of schools are identified in the 'follow-up' section of Figure 3:

- S1 schools tested a sample of pupils for the primary outcome measures;
- S2 schools tested the entire cohort for the primary outcome measures; and
- S3 schools withdrew from the primary outcome measures.

---

<sup>17</sup> One school withdrew participation without the knowledge of group allocation and another school was randomised due to an administrative error.

In addition to distinguishing three types of schools, it is also important to identify different types of pupils based on whether they had any test data.

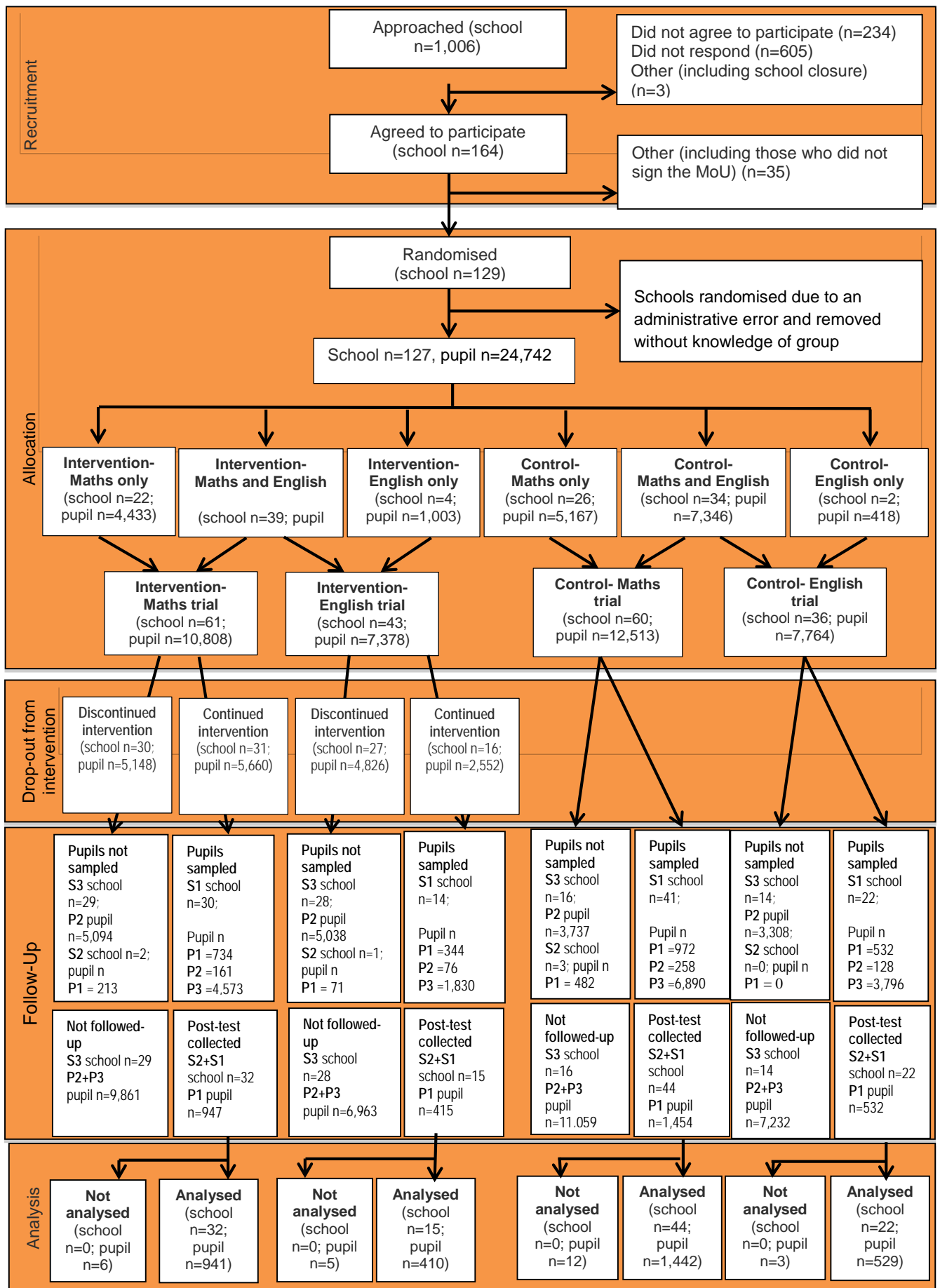
- P1 pupils had the data on the primary outcome measures. They were from S1 schools where they were sampled to take the tests or they were from S2 schools where the entire cohort was being tested.
- P2 pupils did not have any data on the primary outcome measures. These pupils were either sampled to be tested (S1 schools) or belonged to S2 schools where the entire cohort was being tested, or belonged to S3 schools that withdrew from the tests.
- P3 pupils were not followed up and therefore did not have any data on the primary outcome measures. These pupils were non-sampled pupils from S1 schools and therefore no test data was expected for them.

Above combination of types of schools and types of pupils are presented in the 'follow-up' section of Figure 3.

The 'Analysis' section in Figure 3 presents the number of cases that were included, and excluded, from the analysis. A handful of pupils were not included in the analysis (18 pupils for the maths trial and eight pupils for the English trial respectively) as they had outcome measures but their prior attainment data was missing from the NPD.

Table 8 provides details of minimum detectable effect sizes at different stages in the trial.

Figure 3: Participant flow diagram for BPS trial



S1-Schools tested a sample of pupils for the primary outcome measures

S2-Schools that tested the entire cohort for the primary outcome measures

S3-Schools withdrew from the primary outcome measures

P1-Pupils with the data on primary outcome measures

P2-Pupils (who were meant to be) but did not have the data on primary outcome measures

P3-Pupils not sampled and (were not meant to be followed up) for the primary outcome measures



**Table 8: Minimum detectable effect size at different stages**

	Stage	n schools (n=each group) n pupils (n=each group)	Correlation between pre- test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Maths Trial	Protocol (maths trial)	120 schools (60 intervention; 60 control) 3,600 pupils (1,800 intervention; 1,800 control)	0.70	0.15	Five blocks, stratified by setting practice (English/mat hs/both)	80%	0.05	0.16
	Randomisati on (maths trial)	121 schools (61 intervention; 60 control) 3,630 pupils <sup>18</sup> (1,830 intervention; 1,800 control)	0.70	0.15	Five blocks, stratified by setting practice (English/mat hs/both)	80%	0.05	0.16
	Analysis (i.e. available pre- and post- test) (maths trial)	76 schools (32 intervention; 44 control) 2,383 pupils (941 intervention, 1,442 control)	0.78	0.10		80%	0.05	0.15
English Trial	Protocol (English trial)	120 schools (60 intervention; 60 control) 3,600 pupils (1,800 intervention; 1,800 control)	0.70	0.15	Five blocks, stratified by setting practice (English/mat hs/both)	80%	0.05	0.16
	Randomisati on (English trial)	79 schools (43 intervention; 36 control) 2,370 pupils (1,290 pupil; 1,080 pupils)	0.70	0.15	Five blocks, stratified by setting practice (English/mat hs/both)	80%	0.05	0.20
	Analysis (i.e. available pre- and post-test) (English trial)	37 schools (15 intervention, 22 control) 939 pupils (410 intervention, 529 control)	0.66	0.03		80%	0.05	0.19

For further clarity, Table 9 shows the number of pupils sampled for post-testing, the number of pupils analysed and the percentage analysed out of the pupils sampled.

<sup>18</sup> Note: At the time of randomisation, we did not have pupil list from the participating schools. Therefore, these are estimated pupil numbers. These are different from the numbers presented in Figure 3 which reflects the pupil numbers based on the list provided by the schools and the NPD.

**Table 9. Pupils analysed out of those sampled for post-testing**

	Maths			English		
	Intervention	Control	Total	Intervention	Control	Total
<b>Pupils sampled for post-testing</b>	1,825	1,800	<b>3,625</b>	1,290	1,080	<b>2,370</b>
<b>Analysed (no.)</b>	941	1,442	<b>2,383</b>	410	529	<b>939</b>
<b>Analysed (%)</b>	52%	80%	<b>66%</b>	32%	49%	<b>40%</b>

Note: As described in the section on ‘participants’, some schools tested a sample of pupils (S1), other schools tested their entire Year 8 cohort (S2) and the rest of the schools withdrew from the primary outcome tests (S3). We didn’t sample any pupils in the last two types of schools (S2 and S3). The total number of pupils ‘available for sampling’ are presented in Figure 3 to present the actual flow of the participants and how they were included or excluded from the analysis, whereas the above table presents the number of pupils sampled for post-testing *if S2 and S3 schools had allowed us to sample these pupils for post-testing*. These numbers are similar to the randomisation numbers presented in Table 8. The only slight difference arises due to not enough pupils being available for sampling for schools in the intervention group for the Maths trial.

### Pupil characteristics

Tables 10 and 11 present baseline characteristics of the schools that took part in the maths and English trials respectively. The difference in prior attainment between intervention and control pupils (for the analysed groups), expressed as an effect size from a multilevel model for maths, was -0.05 (-0.20, 0.09) and for English -0.01 (-0.18, 0.16). Further exploration of bias that might have ensued through attrition was carried out through using logistic models of the missingness mechanism (see ‘Missing data and sensitivity analysis’).

Table 10: Baseline comparison, Maths trial

Variable	Intervention group		Control group	
School-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
<b>School Governance:</b>				
Academy or Free School	17/32 (0)	53%	28/44 (0)	64%
Maintained	15/32 (0)	47%	16/44 (0)	36%
<b>Ofsted rating:</b>				
Outstanding	9/31 (1)	29%	11/43 (1)	26%
Good	18/31 (1)	58%	24/43 (1)	56%
Requires improvement	3/31 (1)	10%	6/43 (1)	14%
Inadequate	1/31 (1)	3%	2/43 (1)	5%
Urban	29/32 (0)	91%	36/44 (0)	82%
Rural	3/32 (0)	9%	8/44 (0)	18%
<b>Secondary School Type:</b>				
Comprehensive to 16	12/30 (2)	40%	13/44 (0)	30%
Comprehensive to 18	18/30 (2)	60%	31/44 (0)	71%
<b>Percentage pupils eligible for FSM 2015/2016 (5 point scale):</b>				
Lowest 20%	3/32 (0)	9%	7/44 (0)	16%
2nd lowest 20%	7/32 (0)	22%	14/44 (0)	32%
Middle 20%	8/32 (0)	25%	7/44 (0)	16%
2nd highest 20%	12/32 (0)	38%	12/44 (0)	27%
Highest 20%	2/32 (0)	6%	4/44 (0)	9%
<b>School-level (continuous)</b>	<b>n (missing)</b>	<b>[Mean]</b>	<b>n (missing)</b>	<b>[Mean]</b>
N12 Number of pupils whose 12 <sup>th</sup> birthday is in 2015/2016	31 (1)	182	44 (0)	201
<b>School-level (baseline grouping practices, categorical)</b>	<b>n/N (missing)</b>	<b>Percentage</b>	<b>n/N (missing)</b>	<b>Percentage</b>
<b>Set in Year 7 maths?</b>				
Yes	28/32 (0)	88%	42/43 (1)	98%
No	4/32 (0)	13%	1/43 (1)	2%
<b>Stream in Year 7 maths?</b>				
Yes	6/32 (0)	19%	3/43 (1)	7%
No	26/32 (0)	81%	40/43 (1)	93%
<b>Set in Year 8 maths?</b>				
Yes	29/32 (0)	91%	43/43(1)	100%
No	3/32 (0)	9%	0/43(1)	0%
<b>Stream in Year 8 maths?</b>				
Yes	4/32 (0)	12%	2/43 (1)	5%
No	28/32 (0)	88%	41/43 (1)	95%
<b>Student-level (continuous)</b>	<b>n (missing)</b>	<b>[Mean]</b>	<b>n (missing)</b>	<b>[Mean]</b>
% eligible for FSM (Ever, Spring 2015)	939 (2)	27%	1,441 (1)	32%
KS2 maths point score	941 (0)	29.16	1,442 (0)	29.63

Table 11: Baseline comparison, English trial

Variable	Intervention group		Control group	
School-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
<b>School Governance:</b> Academy or Free School Maintained	9/15 (0) 6/15 (0)	60% 40%	16/22 (0) 6/22 (0)	73% 27%
<b>Ofsted rating:</b> Outstanding Good Requires improvement Inadequate	2/14 (1) 10/14 (1) 2/14 (1) 0/14(1)	14% 71% 14% 0%	5/22 (0) 12/22 (0) 4/22 (0) 1/22 (0)	23% 55% 18% 5%
Urban Rural	14/15 (0) 1/15 (0)	93% 7%	18/22 (0) 4/22 (0)	82% 18%
<b>Secondary School Type:</b> Comprehensive to 16 Comprehensive to 18	8/15 (0) 7/15 (0)	53% 47%	9/22 (0) 13/22 (0)	41% 59%
<b>Percentage pupils eligible for FSM 2015/2016 (5 pt scale)</b> Lowest 20% 2nd lowest 20% Middle 20% 2nd highest 20% Highest 20%	0/15 (0) 2/15 (0) 5/15 (0) 7/15 (0) 1/15 (0)	0% 13% 33% 47% 7%	1/22 (0) 11/22 (0) 5/22 (0) 4/22 (0) 1/22 (0)	5% 50% 23% 18% 5%
<b>School-level (continuous)</b>	<b>n (missing)</b>	<b>[Mean]</b>	<b>n (missing)</b>	<b>[Mean]</b>
N12 Number of pupils whose 12 <sup>th</sup> birthday is in 2015-16	14 (1)	161	22 (0)	204
<b>School-level (baseline grouping practices, categorical)</b>	<b>n/N (missing)</b>	<b>Percentage</b>	<b>n/N (missing)</b>	<b>Percentage</b>
<b>Set in Year 7 English?</b> Yes No	13/15(0) 2/15(0)	87% 13%	20/22 (0) 2/22 (0)	91% 9%
<b>Stream in Year 7 English?</b> Yes No	2/15(0) 13/15(0)	13% 87%	3/22 (0) 19/22 (0)	17% 86%
<b>Set in Year 8 English?</b> Yes No	12/15(0) 3/15(0)	100% 0%	22/22 (0) 0/22 (0)	100% 0%
<b>Stream in Year 8 English?</b> Yes No	2/15(0) 13/15(0)	13% 87%	2/22 (0) 20/22 (0)	9% 91%
<b>Student-level (continuous)</b>	<b>n (missing)</b>	<b>[Mean]</b>	<b>n (missing)</b>	<b>[Mean]</b>
% eligible for FSM (Ever, Spring 2015)	409 (1)	32%	529 (0)	25%
KS2 reading point score	410 (0)	29.33	529 (0)	29.44

## Outcomes and analysis

Table 12 presents the results from the primary ITT analysis of the impact of the intervention on pupil outcomes for both subjects. The effect sizes for the impact of the intervention on Year 8 pupils PTM13 and PTE13 scores were -0.01 (-0.11, 0.10) and -0.08 (-0.21, 0.05) respectively. These differences were not statistically significant at the 5% level and suggest that the small differences between the groups are likely to be due to chance. As these models included the variable that was used to stratify the randomisation, we also ran another pair of models so that they can be included in the EEF's cross-study analysis. For PTM13, the effect size was -0.01 (-0.12, 0.10). For PTE13, the effect size was -0.08 (-0.22, 0.05). We also ran models including one school that tested Year 8 pupils using the online PTM13 and PTE13 tests. Although these models show an incremental change in the coefficients for the intervention group, the results show that these were not statistically significant. Therefore, they suggest there is no evidence of impact of the intervention.

**Table 12: Results from the primary and secondary analyses**

Outcome	Raw means				Effect size		
	Intervention group		Control group		n in model (intervention; control)	Hedges g equivalent (95% CI)	p- value
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)			
<b>Maths</b>	32 schools; 941 students (29 schools)	30.83 (29.88, 31.77)	44 schools; 1,442 students (16 schools)	31.10 (30.33, 31.87)	2,383 (941; 1,442)	-0.01 (-0.11, 0.10)	0.91
<b>English</b>	15 schools; 410 students (28 schools)	30.18 (28.92, 31.44)	22 schools; 529 students (14 schools)	31.69 (30.51, 32.88)	939 (410; 529)	-0.08 (-0.21, 0.05)	0.23
<b>Maths (everFSM only)</b>	31 schools; 249 students (30 schools)	26.25 (24.43, 28.07)	44 schools; 464 students (16 schools)	26.00 (24.61, 27.38)	713 (249; 464)	0.02 (-0.14, 0.18)	0.80
<b>English (everFSM only)</b>	15 schools; 131 students (28 schools)	27.80 (25.64, 29.96)	22 schools; 134 students (14 schools)	27.28 (25.01, 29.54)	265 (131; 134)	-0.01 (-0.25, 0.22)	0.90
<b>Maths self- confidence</b>	27 schools; 2,445 students (34 schools)	3.64 (3.61, 3.68)	33 schools; 3,584 students (27 schools)	3.59 (3.56, 3.63)	6,029 (2,445; 3,584)	0.02 (-0.08, 0.13)	0.66
<b>English self- confidence</b>	13 schools; 972 students (30 schools)	3.58 (3.52, 3.63)	17 schools; 2,074 students (19 schools)	3.45 (3.41, 3.49)	3,046 (972; 2,074)	0.09 (-0.08, 0.25)	0.29

### Missing data and sensitivity analysis

For schools that participated in outcome testing, a random sample of 30 pupils was drawn from each school (stratified by FSM eligibility using everFSM variable from the NPD) for English, mathematics, or both according to each school's participation as detailed in the protocol. Outcome data was considered missing if a pupil was sampled but did not have an outcome measure (because they were absent on the day of test or left the school); non-sampled pupils were not considered to be missing, as it was never intended to collect their data. Where schools tested their whole Year 8 cohort—that is, where no sampling took place—pupils who were either without results, or whose results could not be matched to their background data, were considered missing.

For schools that did not complete/participate in the outcome testing, outcome data for all their pupils was missing. For these schools, the sampling of 30 pupils per subject (stratified by everFSM eligibility) was replicated (similar to the schools that took part in the outcomes testing), and these sampled pupils were included in the missing data analysis; non-sampled pupils were not.

### Mathematics results

In the mathematics trial we had 2,401 pupils with PTM13 mathematics outcome scores (18 of which had missing KS2 scores), and 1,976 pupils who should have had PTM13 mathematics outcome scores but were missing.

The probability that the outcome measure was observed (compared to missing) was modelled using a multilevel logistic model and was found to be significantly associated with (a) being in the control group, (b) whether the school took part in one subject or both, (c) KS2 mathematics point score, (d) pupils' everFSM eligibility, and (e) schools' setting practice in Year 7. This shows that the data is not missing completely at random (MCAR).

Missing data was imputed (with chained equations, implemented using the MICE package in R) under the assumption that data was missing at random (MAR). The mathematics model was a multilevel normal regression model for continuous variables and a single-level logistic regression model including school IDs as fixed effects for binary variables. The imputation model included variables PTM13 mathematics raw score, KS2 mathematics point score, gender, everFSM eligibility, intervention or control group, an indicator of whether the school took part in one subject or both, and whether the school set pupils in Year 7 before the trial.

The main ITT model was run using each of the imputed datasets and the results were pooled to give coefficients and standard errors that take account of the imputation variance. This complete data analysis gave the coefficient of being in the interventions as -0.28 (-3.40, 2.83). This compares to a completers model raw intervention coefficient of -0.09 (-1.66, 1.45).

To explore further the assumption that data is missing not at random we adjusted the values of PTM13 Mathematics raw score in the imputed datasets in a sensitivity analysis. The standard deviation of PTM13 Mathematics raw scores is 14.9 so we adjusted by values ( $\delta$ ) ranging from +/- 7.5 (about half a standard deviation) and observed the impact on the imputed datasets. The SAP described  $\delta$  in standard deviation units (-0.5 to 0.5) so we converted it to raw units for the analysis.

**Table 13: Results from the sensitivity analysis for maths primary outcome measure**

$\delta$	coefficient for intervention	s.e.
-7.5	-1.28	1.39
-5	-0.62	1.54
-3	-1.13	1.67
-1.5	-0.63	1.48
0	0.19	1.16
1.5	-0.10	1.67
3	0.23	1.55
5	0.39	1.51
7.5	0.53	1.46

This sensitivity analysis shows that as  $\delta$  increases the coefficient for intervention tends to increase but is somewhat erratic. This would suggest that the results of our multiple imputation are sensitive to variations in the data that are not accounted for in the imputation model. However, it should be noted that even with very low and high values of  $\delta$ , the coefficient remained lower than its standard error. This implies that even with quite extreme assumptions regarding the values of missing data, the result remained well within the bounds of chance rather than genuine change.

### English results

In the English trial we had 947 pupils with PTE13 English outcome scores (eight of which had missing KS2 scores), and 1,483 pupils who should have had PTE13 English outcome scores but were missing.

The probability that the outcome measure was observed (compared to missing) was modelled using a multilevel logistic model and was found to be significantly associated to (a) being in the control group, (b) KS2 reading point score, (c) pupils' everFSM eligibility and (d) schools' setting practice in Year 7. This shows that the data is not missing completely at random (MCAR).

Missing data was imputed (with chained equations, implemented using the MICE package in R), under the assumption that data was missing at random (MAR). The English model was a single-level<sup>19</sup> normal regression model for continuous variables and a single-level logistic regression model, including school IDs as fixed effects for binary variables. The imputation model included variables PTE13 English raw score, KS2 reading point score, gender, everFSM eligibility, intervention or control group, an indicator of whether the school took part in one subject or both, and whether the school set pupils in Year 7 before the trial.

The main ITT model was run using each of the imputed datasets, and the results were pooled to give coefficients and standard errors that take account of the imputation variance. This complete data analysis gave the coefficient of being in the intervention as -1.01 (-2.38, 0.36). This compares to a raw coefficient in the English completers model of -1.09 (-2.82, 0.65).

The standard error of the intervention coefficient was smaller in the English complete data model than in the mathematics one below, despite there being more missing data in the English model than in the

<sup>19</sup> The extent of missing data, 61%, meant that a multilevel model was not implemented due to convergence problems.

mathematics model. This is probably due to being unable to account for the clustering of the data in the English imputation model (it was single-level).

To explore further the assumption that data is missing at random we adjusted the values of PTE13 English raw score in the imputed datasets in a sensitivity analysis. The standard deviation of PTE13 English raw scores is 13.5, so we adjusted by values ( $\delta$ ) ranging from +/- 7 (just over half a standard deviation) and observed the impact on the imputed datasets. The SAP described  $\delta$  in standard deviation units (-0.5 to 0.5) so we converted it to raw units for the analysis.

**Table 14: Results from the sensitivity analysis for English primary outcome measure**

$\delta$	coefficient for intervention	s.e.
-7	-2.46	1.00
-5	-1.99	0.84
-3	-1.58	0.80
-1.5	-1.03	0.79
0	-0.99	0.75
1.5	-0.60	0.76
3	-0.54	0.74
5	-0.16	0.82
7	0.59	0.95

This sensitivity analysis shows that as  $\delta$  increases the coefficient for intervention increases. This would suggest that the results of our multiple imputation are sensitive to variations in the data that are not accounted for in the imputation model. This further illustrates how unreliable the result is. The situation for English was even more volatile than for maths due to the greater proportion of missing data. Here, the intervention coefficient became significantly negative for highly negative values of  $\delta$  whilst never becoming significantly positive. These results demonstrate that we are marginally less sure of the null effect than we are for maths.

### Subgroup analyses

We also conducted analyses for everFSM pupils only. These models were similar to those of the ITT models. As seen in Table 12, results from these models suggest that the intervention has no effect on any of the outcomes for everFSM pupils. The effect size for this subgroup analysis was 0.02 for the maths outcome and -0.01 for English outcome and suggest no evidence of impact of the intervention on attainment of everFSM pupils.

In addition, we also conducted further analysis to measure the differential impact of the intervention on everFSM pupils compared to those who were not everFSM. For the maths outcome, raw coefficient of the interaction term was 0.93 (se=0.85, p=0.27) and for the English outcome, it was 0.79 (se=1.46, p=0.59). This suggests that the results were not statistically significant and there is no evidence of differential impact when pupil everFSM eligibility is considered.

We also ran models to measure the differential impact of the intervention based on pupil prior attainment. The raw coefficient for the interaction term for the maths outcome was -0.1 (se=0.08, p=0.21) and was -0.1 (se=0.15, p=0.52) for the English outcome, suggesting that there is no evidence of differential impact of the intervention given pupil prior attainment.



## On-treatment analysis

UCL Institute of Education collected fidelity data throughout the two-year delivery period. The measures and methods of data collection were pre-agreed between the two organisations. Fidelity measures related to the professional development sessions and were based on schools' attendance at these sessions and therefore had good data coverage. Other fidelity measures were dependent on schools sending their data to UCL Institute of Education. Despite numerous efforts to obtain this information from the schools, a large number did not return it to UCL Institute of Education. Over time, this led to a large amount of missing data.

The Statistical Analysis Plan specified that the relationship between fidelity and outcomes would be measured using composite measure 2 (see the methods section). Unfortunately, this measure was available for only 11 intervention schools for the maths trial (out of 62) and three intervention schools in the English trial (out of 43). This limited data was not suitable for quantitative analysis and we therefore decided to present this data within the wider qualitative findings. This is presented in the process and implementation findings in the 'Fidelity' section.

## Secondary outcomes analyses—self-confidence measures

We also ran completers models for pupil self-confidence outcomes. Results from these models are presented in Table 12. The maths self-confidence model included all pupils with both baseline and end-point maths self-confidence measures from schools that took part in the maths trial. The effect size for this analysis was 0.02 (-0.08, 0.13). The English self-confidence model included all pupils with both baseline and end-point English self-confidence measures from schools that took part in the English trial. The effect size for this analysis was 0.09 (-0.08, 0.25). Results from both these models suggest that there is no evidence of impact of the intervention on pupil self-confidence for either subject.

## Cost analysis

### How much does it cost?

The intervention cost a total of £3,648 per department per school or £20.04 per student for the first year of implementation, with the main financial costs being the developers' staff and monetary costs (such as staff travel and venue hire) for delivering the training, together with teacher participants' travel and subsistence and printing and photocopying costs. The cost of training and travel was covered by the EEF as part of this trial. In addition, participating schools reported using an average of two hours of cover per department per year so that staff could take part in the initiative. The annual cost for the second year of implementation was £4,016 per department or £22.07 per student per department. The average costs per department per year were higher in the second year owing mainly to the fact that UCL staff delivery costs were slightly higher in 2016/2017, while the number of participating English and maths departments was lower (due to attrition), resulting in a higher cost per department per year. As this was a two year intervention, and the training was cascaded to staff across participating departments, we have assumed there would be no additional costs to schools if it was continued into a third year. The cumulative cost per student per year equates to an EEF cost rating of 'very low'.

The cost of the intervention for this evaluation, presented below, has been estimated on the basis of costs that were part of the project. While it provides a rough estimate of the cost that might be expected, the costs do not necessarily represent what the cost would be to a school if the intervention was scaled-up, for example, if the staff delivering the intervention changed.

This cost evaluation estimates the cost to schools under the assumption that the EEF funding for financial costs, prerequisite costs and compensation for staff time is not being provided and that schools are paying for their share of the total costs they would otherwise bear. As the intervention requires

schools to modify their existing practice, we have worked on the assumption that there were no additional prerequisite costs.

### Financial costs

The single main financial cost of delivering the intervention, which was borne by the EEF as part of this project, was that relating to UCL staff costs for delivering training, which equated to costs of £252,262.00 in 2015/2016, and £260,280.00 in 2016/2017. Other direct, marginal costs, such as staff travel costs and venue hire, were also recorded and are presented in Table 15 below. This information was kindly supplied by UCL on a termly basis. As the same UCL delivery team were working on both the Best Practice in Setting efficacy trial, and the Mixed Ability Grouping pilot study, this information was aggregated across both interventions. However, it has been possible to split direct marginal costs across both interventions using the ratio 0.983 (for the Best Practice in Setting initiative) and 0.017 (for the Mixed Ability Grouping pilot study). These ratios represent the relevant award of funds for the two interventions from the EEF grant, and have been agreed with the developer. It has not been possible to provide a similar breakdown to the staff costs as the delivery team have recorded their time as working on one holistic project, rather than as standalone contributions to two separate interventions. The staff costs for delivering the Best Practice in Setting initiative are therefore overinflated, as the costs include time required to deliver the Mixed Ability Grouping pilot study. To calculate a more accurate cost per department per year in the cost per student per year calculations, the UCL staff costs were divided by the total number of departments participating in both the Best Practice in Setting initiative and the Mixed Ability Grouping pilot study.

**Table 15: UCL’s financial costs for delivering the Best Practice in Setting initiative and the Mixed Ability Grouping pilot study**

	2015/2016	2016/2017
UCL Staff travel	£2,214.89	£1,329.65
UCL Staff hospitality	£742.91	£244.87
Teacher travel and subsistence	£2,635.80	£332.89
Professional development venue and catering	£6,617.80	£528.85
Professional development additional delivery support	£614.38	£1,213.51
Print materials	£4,933.68	£0.00
Web materials	£1,503.87	£0.00
<b>Total</b>	<b>£19,263.32</b>	<b>£3,649.77</b>

We also asked schools whether they had incurred additional costs as part of their implementation of the intervention. We received responses from 16 schools to these proforma questions in 2015/2016 (equating to 25% of the total number of intervention schools),<sup>20</sup> and responses from 20 schools in 2016/2017 (equating to 31% of the total number of intervention schools). Schools’ responses indicated that little additional financial expenditure was required, particularly in the second year of implementation. Using closed response options, those that responded cited the costs of:

<sup>20</sup> Note that this is not a response rate as some intervention schools had already withdrawn from the trial.

- printing/photocopy or material costs, which equated to an average of £35 per department in 2015/2016 and £0 in 2016/2017; and
- travel and subsistence costs (related to CPD, not delivered by UCL), which equated to an average per department cost of £40.55 in 2015/2016, and £6.05 in 2016/2017.

We include these figures in our estimate of total cost, although it is a rough estimate, given the limited information it is based on.

UCL collected information about when the schools dropped out from the intervention. This was collected and recorded between the professional development sessions. Using this, we calculated the number of schools that were still continuing to participate in each intervention in each academic year to provide, first, average costs per department per year, and, second, average costs per student per department per year. Table 16 shows our estimate of the cost per student per department (English or mathematics) over time, the cumulative cost per student over three years, and the average cost per student per year. We derived the cost per student by dividing the average cost per department per school per year by the average number of Year 7 students per school (using student administrative data collected by NFER). This calculation yields a total cost per student over three years of £42.11 and an average cost per student per year of £14.04. The cumulative cost per student per year equates to an EEF cost rating of 'very low'.

**Table 16: Average financial cost per student per year**

Cost item	Cost per student	Cumulative cost per student	Average cost per student per year
First year	£20.04	£20.04	£20.04
Second year	£22.07	£42.11	£21.05
Third year	£0.00	£42.11	£14.04

### Staff time

The Best Practice in Setting initiative involved staff attending four regional workshops across the two-year intervention, with the expectation that approximately four staff from each participating school would attend each workshop. It was expected that one of these would be the relevant Head of Department who would become the agent responsible for implementation, monitoring and working with teachers in their school. Data supplied by the developer suggested that each workshop lasted for approximately two hours. Data was not collected on participants' average travel time to each session. As the training was undertaken after school, supply cover was not required. However, excluding participation in the professional development sessions run by UCL, participating schools reported using an average of two hours of cover per department per year for the first two years of the intervention so that staff could take part in the initiative. Of this time, approximately one hour on average per department per year, required paid supply cover to be used.

## Process evaluation

This section presents the key findings from the light-touch process evaluation. It draws on the findings from telephone interviews with 11 participating teachers, comprising five Heads of English, four Heads of Mathematics, and two school senior leaders (a Deputy Headteacher and a Vice Principal), undertaken between March and June 2016 (the second year of the evaluation). In the third year of the evaluation, a total of five interviews were undertaken (some involving more than one person) between March and May 2017, involving four Heads of English, one Head of Mathematics, one Higher Level Teaching Assistant (HLTA) who spoke about the mathematics department's involvement, and one school senior leader. Five of the interviewees (three from an English department, one from a mathematics department and one senior school leader) were interviewed in both years of the evaluation, see Table 17 below. Each interview lasted for approximately 30 minutes. It should be noted that this is a small sample size and as such caution should be taken when considering over-arching conclusions. However, schools provided useful reflections that would be helpful for anyone considering implementing a similar intervention in the future.

**Table 17: Process interviews undertaken in each year of the evaluation**

School	Year 2 of evaluation (2015/2016)		Year 3 of evaluation (2016/2017)	
	Role of interviewee(s)	Department(s) represented	Role of interviewee(s)	Department(s) represented
1	Head of English Head of Mathematics	English Mathematics	Head of English** HLTA	English** Mathematics
2	Senior leader	English Mathematics	-	-
3	Head of English	English	-	-
4	Head of Mathematics	Mathematics	-	-
5	Head of English Head of Mathematics	English Mathematics	-	-
6	Head of English Head of Mathematics	English Mathematics	-	-
7	Senior leader	English Mathematics	Senior leader Head of English Head of Mathematics	English Mathematics
8	Head of English	English	Head of English	English
9	-	-	Head of English	English

*\*\*This interviewee explained that the English department had withdrawn from the intervention in September 2016 (Year 3 of the evaluation) but was still able to talk about the intervention based on the previous 12 months.*

The findings were analysed (and are reported) thematically.

## Implementation

**Most interviewees reported that, overall, their staff had responded positively to the initiative, although a minority of respondents were frustrated with the limitations placed on the movement of students between sets.**

Respondents from most schools (n=7) reported that staff in their departments had generally responded positively to the initiative. This was the case in both English and mathematics departments. However, some suggested that this was due to staff perceiving that they were not being asked to do anything radically different to their usual practice. Indeed, some staff were reportedly 'indifferent' or even 'oblivious' to the intervention. This response appears to reflect the view, held by many schools, that they were already implementing the best practice principles to setting prior to starting the intervention. It should be pointed out that while the fidelity measures show poor implementation of some features (e.g. teacher randomisation), perhaps suggesting a misunderstanding about what the intervention entailed, we found no widespread evidence of this. Instead, the fact that some schools had only partially randomised teachers to pupil groups was explained by schools as being the result of their efforts to make the principles more implementable.

In only one school did the Head of English report that his staff had, collectively, responded negatively to the intervention. This was attributed to a range of factors, including the perception that the venue for the regional workshop was too far away, and frustration with the requirement that students could be moved between sets only three times across the two-year period. The last point was a source of frustration for staff in three schools, who would have preferred greater freedom over when they moved students.

## Supporting factors

Interviewees identified a number of supporting factors associated with their approach to implementation that they believed helped the intervention to run smoothly. These included:

- the school already having an ethos/culture that set high expectations for all students;
- staff being open-minded towards, and embracing of, the intervention;
- re-labelling attainment sets in a way that did not attach a 'high', 'medium' or 'low' designation, such as moving away from a numbering system (students were reported to be more positive about being placed in different sets as a result); and
- having good parental support.

## Challenges

In response to a direct question about the challenges associated with implementation, two interviewees identified the following issues:

- it was difficult not to overburden staff with too much information; and
- staff who were now teaching different attainment groups to those that they were accustomed to required support.

In addition, while most interviewees reported that their departments had followed the developer's guidance on setting, several interviewees raised concerns about the use of Key Stage 2 results as the basis for allocating a student to a particular attainment set. Several respondents reported that the Key Stage 2 data did not always fit with later assessments that they undertook, such as through teacher assessments or the cognitive abilities test (CAT).

One interviewee reported that they had to 'pacify' some parents when they were informed about their child's allocation, while another said that timetabling had been difficult where students had been placed in different attainment sets for both English and mathematics. It was unclear whether these events would happen normally, or whether they were new developments resulting from the good practice principles underpinning the intervention.

Other challenges associated with implementation of the intervention surfaced during the interviews and are discussed in the sections below.

**Most interviewees reported that the intervention was similar to their usual approaches to setting.**

Interviewees were asked how similar or different the intervention was to their usual approach to setting. Respondents from all nine schools reported that it was very similar, although further questioning often suggested a number of notable differences, some of which are associated with poor prior approaches to attainment setting. It therefore seems that although schools felt they were implementing many of the principles, this may have been due to a misunderstanding of the intervention requirements. For example, some schools reported they would normally:

- move students between sets more or less frequently than that required by the intervention, while some would do this for reasons other than attainment, for example, by considering pastoral care needs, or to ensure a certain ethnic or gender mix in each set; and
- not randomise the allocation of teachers to different sets: typically other factors would be considered, such as teacher preferences and perceptions of teacher strengths (for example, some teachers were considered better suited to teaching students in high or low sets).

Just one interviewee, a Head of English, reported that her department had been using mixed-ability groupings for their Year 7 students prior to starting the intervention. The mathematics department in the same school was already setting students by attainment.

**Most interviewees were not very positive about the workshops, although many appeared to enjoy the opportunities to network with teachers from other schools.**

Most interviewees appeared to find the opportunities to interact with teachers from other schools, to exchange information, and to develop professional or social contacts to be the most valuable part of the training events. This was particularly the case where staff were teaching the same subject but in different schools. However, not every workshop was reported to be well attended, and this limited the opportunities for discussion.

A number of specific issues were raised about the workshops, which will hopefully prove to be useful for others considering similar interventions, although caution should be applied to findings as each point is raised by only one or two interviewees. To help with considerations about scale-up, these have been separated into those concerning logistical issues (which might be more straightforward to address) and those regarding content and delivery (which might be more challenging to address).

### Logistical issues

Two interviewees reported they would have preferred the workshops to have run in school time, instead of as twilight sessions, as this caused problems with their childcare arrangements and/or with their planning or marking work, which they usually undertook in the evening. However, it is not clear whether schools would have been able to release teachers had the training been delivered during the school day.

One interviewee said the workshop venue was a long way to travel from his school, and each workshop took up to four hours of an evening.

In two cases, interviewees said they had taken part in Skype video conferences, either as an alternative to attending a face-to-face workshop, or to catch up with the developers if they had missed a workshop. In both cases, these were reported to have worked well, and, because of their added convenience, were preferred to the face-to-face meetings. It was unclear from the interviews whether the teacher-to-teacher conversations, which so many interviewees appeared to enjoy the most, could be facilitated through the video conference meetings.

### Content and delivery

One Head of Mathematics reported that there was too much pre-reading before each workshop: 'It was great that information was sent to us in advance, but there was too much of it. It wasn't realistic to expect us to read all that.'

There appeared to be a general belief that the training could have been better tailored to the experiences of those attending. For example, three interviewees reported that the workshops covered material that they already knew very well, such as growth mindsets, and that they did not need additional training on this. Four interviewees suggested that the workshops should have allowed more time for discussions of a practical nature, rather than focusing on the theory behind the intervention.

### Fidelity

The process evaluation explored the extent to which the intervention had been delivered as intended by the developers. The following section discusses each fidelity measure in turn, drawing on the findings from the telephone interviews with participants, and secondary data that was collected or processed by UCL Institute of Education.

#### Attendance at training

**While most participating English and mathematics departments were represented at each training workshop, overall, fewer teachers attended the workshops than was expected.**

Four regional workshops were provided across the two-year intervention. The protocol states that 'approximately 6–10 staff from each school' were expected to attend each training session. However, by the time the MOU was written, this guidance had been revised to suggest that 'two nominated Year 7 English teachers, and two nominated Year 7 mathematics teachers' should attend each workshop. The preference was that one of these teachers should also be the relevant Head of Department (see Appendix C). On the basis of this level of attendance, most of the schools we spoke to appear to have complied fully with the guidance. In a small number of cases, a school senior leader also attended the training, particularly the first workshop, while in at least two schools, only one member of staff attended each workshop. Smaller subject departments said that finding someone to attend the workshops could be challenging, particularly when accounting for staff sick leave and maternity absences.

Data from UCL Institute of Education's Register of Professional Development (see Appendix F) suggested that the expected number and type of staff attending each session declined over the two

years. Attendance from participating mathematics departments was higher than participating English departments for all four professional development sessions.<sup>21</sup> Peak attendance was for session two, which is when most schools had been recruited to the intervention (60% for mathematics; 49% for English), and the lowest recorded attendance was for session four (21% for mathematics; 12% for English).

Drawing on both the interview and UCL data, overall, there appeared to be a low level of fidelity with regard to this measure as although the staff that attended the workshops appeared to be the 'right staff', attendance was poor, particularly towards the end of the intervention.

### **Cascading training**

**Most interviewees reported cascading the training to the other members of their departments, as required by the intervention, although a minority reported doing this on a limited basis only.**

Staff in seven of the nine schools that we spoke to confirmed that all of the training and materials had been shared with other colleagues in their departments. The main method of doing this was through department meetings or twilight continuing professional development (CPD) sessions. Electronic versions of training materials/handouts were circulated to staff by email.

Staff from two of these seven schools reported that while some of the training had been shared, this was undertaken on a limited basis. One school reported they had shared only selected materials as they did not think their staff would find everything of value, while another reported that 'instability' in their staffing had meant that it had been difficult to ensure that new staff were briefed and up-to-speed with the initiative. This last point was a feature of the intervention covering two school years, and highlights the challenges of sustaining the intervention over this time period due to staffing changes.

In contrast to the findings from the interviews, data from UCL Institute of Education suggested that about half of all participating English and mathematics departments were reported to have undertaken some form of information sharing in the earlier stages of the intervention, with this appearing to decline further towards the end of the intervention. However, data was not provided from all schools, so it is not clear what cascading happened in schools that did not attend the training workshops.

Overall, there appeared to be a moderate level of fidelity with regard to this measure.

### **Allocation of students to classes**

**All but one of the schools confirmed that they had adhered to the intervention's requirement to use three or four set levels.**

Most interviewees reported that they were using four, three or two set levels with their Year 7 and 8 students during the evaluation. Just one school reported that its Year 7 and 8 students were grouped into five different set levels. This was reported to be a reduction in the number of set levels the school had used prior to the evaluation, but the school was finding differentiation harder using the smaller number of sets. There were no notable differences in the number of set levels used by English and mathematics departments, and schools where we spoke to representatives from both departments more often than not reported having the same number of sets in each subject area.

Data from UCL Institute of Education suggested that of those 38 schools for which the information had been collected, most schools setting arrangements (n=29) did follow the interventions' best practice

---

<sup>21</sup> A number of schools were unable to arrange for teachers to attend the first round of sessions in July 2015 due to late recruitment, which meant that the final round of randomisation happened on 24 July 2015.



principles. There was, however, a high proportion of missing data (n=23), which makes generalisations to the whole sample difficult.

Overall, there appeared to be a high level of fidelity with regard to this measure.

### **Allocation of teachers to sets**

**Half of the interviewees reported that their teachers had been randomly allocated to sets, as was preferred by the developers, but some reported they did not follow the best practice principles set out by the intervention.**

Of the 14 different English and mathematics departments represented, interviewees representing seven of them reported that their teachers had been randomly allocated to classes, as preferred by the developers. However, data was not gathered on whether it was the school or the developer that undertook this randomisation.<sup>22</sup> Three reported that their teachers had been 'partially randomised', that is to say they had tried to fully randomise the allocation, but other factors had been taken into consideration, such as the hours of part-time teachers, as the following quote illustrates:

*Classes with part-time teachers end up with two teachers. However, we wanted to avoid the bottom set having two teachers, because, in our experience, the students don't cope well with the changes in teaching style, and the randomisation doesn't take account of this.*

Interviewees representing the four remaining departments reported that they had not randomised the allocation of teachers, preferring instead to allocate teachers based on the Heads of Departments' judgements as to which teacher was better at teaching students in different set levels. They were therefore not applying the best-practice principles. Although some constraints on this measure were practical (e.g. to do with teacher availability) others were conceptual, and therefore potentially harder to overcome.

By contrast to the findings reported above, the data from UCL Institute of Education reported that only three schools submitted to independent randomisation as recommended by the developer team. Only two of these schools then applied the randomised teacher allocation, while the rest opted for the 'reserve' option of themselves applying fair principles to allocation. Collectively these findings illustrate a potential lack of understanding between the different parties concerned with the implementation and responses on the part of schools. There was, however, a high proportion of missing data, which makes generalisations to the whole sample difficult.

Overall, there appeared to be a low level of fidelity with regard to this measure, as very few schools were willing to subscribe to the 'optimal' approach of independent teacher randomisation.

### **Allocation of students to sets**

**All but one of the interviewees reported that their students had been allocated to sets based primarily on their Key Stage 2 attainment data, although for some respondents, other factors were also taken into consideration.**

While most interviewees reported using Key Stage 2 attainment data as the primary means of allocating students to different attainment sets, two interviewees reported that this data was supplemented with either teacher assessment data in the first few weeks of Year 7, or with a consideration of students'

---

<sup>22</sup> Where possible, schools were asked to submit to independent randomisation of teachers across sets. Where this was not possible, schools were asked to distribute teachers according to principles provided through the intervention (for example, placing some of their most experienced teachers with lower attaining sets and not assigning NQTs to low sets).

special educational needs (SEN). However, only one interviewee reported that student allocations had been based on a much wider range of factors, including a consideration of friendship groups, their relationships with other staff and students, and their language skills.

Several interviewees reported that Key Stage 2 data was sometimes not available for all students, in which case students were allocated to sets based on teacher assessments.

Data from UCL Institute of Education suggested that of those participating mathematics departments for which we have data (n=38), half had 95% or more of students allocated to classes on the basis of Key Stage 2 results. Of those participating English departments for which we have data (n=22), about a third (n=7) had 95% or more of students allocated to classes on the basis of Key Stage 2 results. There was, however, a high proportion of missing data, which makes generalisations to the whole sample difficult.

Overall, there appeared to be a moderate level of fidelity with regard to this measure.

### **Movement of students between sets**

**Most interviewees reported that students were moved between sets, in line with the intervention's best-practice principles, although there were exceptions.**

Interviewees in seven of the nine schools reported re-setting students in the defined periods:

- January 2016;
- between July and September 2016; and
- January 2017.

In some cases, a small number of students were moved outside of these periods (for example, as a result of behavioural issues or for pastoral reasons), but this was permitted by the intervention, as long as it involved no more than five per cent of the cohort.

Respondents in just two schools reported that student movement between sets was more fluid, with one interviewee arguing that 'the needs of the children have to come before the needs of the trial'.

While, for most interviewees, the initial setting was based on Key Stage 2 attainment data, many interviewees generally reported using a wider range of measures when re-setting, based, for example, on their own assessments and teacher observations as well as feedback from students.

Three respondents reported that they would have liked the option of moving students more frequently, with one reporting that he had received some pressure from parents to move their child. However, in most cases, interviewees did not report experiencing any challenges with complying with this requirement of the intervention.

Data from UCL Institute of Education was inconclusive, owing to the high proportion of missing data, which makes generalisations to the whole sample difficult.

Using the findings from the process evaluation as our primary indicator, overall, there appeared to be a high level of fidelity with regard to this measure.

### **Teachers having high expectations for their students**

**As required by the intervention, all interviewees reported that their teachers had high expectations for their students, and that they had systems in place to ensure that these expectations were maintained.**

These approaches varied, but were often reported to come from a range of measures or features including:

- school culture—senior leaders were seen to play a key role in establishing and maintaining the expectations against which staff and students were judged;
- pedagogy—such as the use of effective differentiation techniques, blind marking and teaching (for example, some teachers were not told who their pupil premium students were, the rationale being that they should seek to achieve the best outcomes for all students), setting aspirational targets for all students regardless of prior attainment, and the use of mastery approaches;<sup>23</sup> and
- recruitment practices—recruiting the most able teachers, including those who could model a Grade 9 GCSE answer, to demonstrate that they could reach the most able students.

While some interviewees acknowledged that some teachers were better at this than others, they all felt that they had processes in place designed to encourage their teachers to have high expectations for all students.

Data from UCL Institute of Education suggested that teacher expectations for students were mixed, with similar proportions of both mathematics and English departments identified as having high expectations for all students as those that did not. One in five teachers in the mathematics departments reported that they had high expectations for all students and similar proportions reported that they did not. These proportions were 16% and 12% respectively for teachers in English departments. There was, however, a high proportion of missing data (60% for maths and 72% for English), which makes generalisations to the whole sample difficult.

Overall, the interview data suggested there appeared to be a high level of fidelity with regard to this measure, although UCL Institute of Education data does suggest expectations may be more mixed than suggested by the interviews alone.

## Outcomes

### Impact on students

Interviewees were asked what impact, if any, the intervention had had on their students.

Of the nine schools that took part in a telephone interview, most (n=6) reported that staff had not noticed any additional positive impacts on their students as a result of participating in the initiative. Most interviewees attributed this to a belief that their schools were already implementing setting according to the best practice principles promoted by the intervention, and that consequently there had been little change to the way they approached attainment grouping. One English department, which withdrew from the evaluation in the third year of the study, reported that one year of participation was not sufficient for them to have been able to observe any positive impacts on their students.

---

<sup>23</sup> An approach whereby teachers spend more time going into depth about a subject as opposed to racing through the things that all children are expected to know.

While some interviewees appeared to find value in the intervention's confirmation that they were approaching attainment grouping based on the latest research evidence, a minority were disappointed with the outcomes of the intervention. One Head of Mathematics felt the focus of the intervention had been unclear and that their school 'got very little from the research'.

Interviewees from two of the nine schools reported that the intervention had led to positive impacts on their staff and students. These impacts included: staff being more proactive in checking that their students were in the most appropriate groups; students feeling 'more settled' in their groups; the pace of lessons being more appropriate to the academic attainment of the group; and staff applying more effective differentiation strategies.

Finally, two interviewees reported that their departments (one English, one mathematics) now placed a greater focus on the rigour and frequency of their in-year testing, and the need to explicitly link this data to the movement of students between attainment groups.

### **Unintended consequences**

Interviewees identified a number of unintended consequences associated with implementation of the intervention, most of which were positive in nature.

In one school, the Heads of English and mathematics both reported that their 'low' and 'top' set Year 7 classes had particularly benefited from the intervention, but that this appeared to have been at the expense of the middle set, who 'lacked the aspiration to do better'. These criticisms appear to focus on the practice of setting, rather than specifically on the best practice principals recommended by the intervention. However, they do perhaps highlight that in ensuring that students with low prior attainment are not disadvantaged by setting, it is important to ensure that students with different levels of prior attainment, and particularly those with middling levels of prior attainment, remain a priority and are not overlooked. Interestingly, for these two departments, the improvements in the performance of students in low and top sets were reported towards the end of the second year of the evaluation, suggesting that impacts on students could be possible after just a few months of the intervention. Unfortunately, we were unable to speak to the same interviewees in the third year of the evaluation and so could not explore whether these impacts were sustained into Year 8. Another unintended consequence, but a positive one, was that the intervention had brought together staff from both the English and mathematics departments. In at least two schools, staff were now sharing and discussing their practice with one another in a way that was not happening before. However, it is not clear how long these new working relationships lasted, with one school reporting that they had not continued into the third year of the evaluation, despite both departments continuing with the trial.

### **Views on scale-up**

Interviewees were asked whether, in its current form, the Best Practice in Setting initiative could be used by other schools, and was therefore suitable for scale-up. Interviewees were divided in their views. Respondents from four schools reported that they thought the intervention was scalable, but that this was subject to participating schools:

- not having other limiting factors, such as high levels of poor student behaviour, or problems with recruiting or retaining high-quality teachers; and
- being committed to the best practice in setting approach and any necessary changes to their current setting practices.

Of the remainder, interviewees from three schools reported they were unsure or did not know, while interviewees from two schools thought the intervention was not yet suitable for scale-up. The reasons for this related to the issues, reported earlier, about implementation and fidelity, such as not all staff

'buying in' to the approach of setting by attainment, frustration with the limitations placed on the movement of students between sets, and the logistical and 'content and delivery' related issues with the workshops.

### **Suggestions for improvement**

A number of suggestions were made for how the intervention could be improved (from both those who thought the intervention was ready for scale-up, as those that did not). These included (in no particular order):

- The workshops should focus less on theory, and more on providing practical guidance on teaching through attainment setting.
- The staff and student surveys should be removed from the intervention as they were perceived as burdensome and difficult to answer (these formed part of UCL Institute of Education's internal evaluation of the intervention and would not normally form part of the intervention).
- The intervention could include more guidance on how schools might evaluate the impact of their setting practices.
- The developers could help to facilitate a 'richer' discussion between teachers participating in the workshops.<sup>24</sup>

---

<sup>24</sup> Note that we did not collect any process evaluation data from the control schools.

## Conclusion

### Key conclusions

1. The project found no evidence that the Best Practice in Setting intervention improves maths or English attainment for children in Years 7 and 8. For English, pupils in the intervention schools made slightly less progress than the control pupils, but this finding has no meaning given its level of uncertainty and very low security rating.
2. There was no evidence of impact on pupils' self-confidence in maths. For English there was a small positive difference in pupil self-confidence, but this result has no meaning due to its level of uncertainty and the large amount of measurement attrition.
3. School and teacher buy-in was low. Half of the schools in the maths trial and more than half of the schools in the English trial ceased intervention delivery before follow-up. Attendance at training sessions decreased over time (to 21% for the final maths session and 12% for the final English session). With this level of treatment attrition, being able to engage schools with the programme was demonstrated to be at least as important as the programme itself. Future interventions aimed at changing setting practices should be designed with the issue of engagement in mind.
4. The process evaluation revealed mixed views from participants. While some were largely positive about the intervention, many interviewees thought that what they were being asked to do represented little change from what they already do. Some schools reported that the intervention was onerous and difficult to deliver.
5. Some schools struggled to implement the intervention in full (for example, there was poor fidelity to teachers being randomly allocated to sets), this seemed to be due to schools finding it difficult to make some changes and therefore adapting principles to make them more implementable.

### Interpretation

The evaluation findings suggest that there is no evidence of impact as neither primary outcome was affected by the intervention. There was a high rate of measurement attrition, which necessitated further work on multiple imputation and sensitivity analysis. Although this revealed considerable volatility in the result, it never passed the threshold of being significantly positive and we can therefore be reasonably confident in the null result. Aside from the intervention itself being ineffective, the null result could be for a combination of two reasons: the lack of implementation fidelity and the trial being underpowered. These are discussed in turn.

There are several key pieces of evidence that suggest that some schools did not 'buy in' to this intervention to the extent that they needed to in order for any impact to be seen:

- Thirty (out of 61) maths intervention schools and 27 (out of 43) English intervention schools dropped out of delivering the programme. Attendance at workshops reduced dramatically during the course of intervention delivery.
- Of those interviewed, a large proportion of teachers thought that what they were being asked to do was no different to what they do normally. This view was, however, contrasted with poor fidelity to some practices, with data suggesting that some elements of the programme (for example random allocation of teachers to sets) were not implemented by the majority of schools. This seems to be due to schools adapting principles to make them more implementable. In its current form the programme does not result in schools making the prescribed adaptations to setting practices.

It is therefore difficult to conclude that the intervention itself is ineffective; we can say merely that it was not suitable for schools in its present form, although some teachers did recommend that it was suitable for other schools in its present form. Had implementation fidelity been better, there might still have been doubt surrounding a null result. This was an efficacy trial with no existing data to inform estimation of an effect size for sample size calculations. Given that the trial was designed for schools that already use setting, the intervention was merely intended to tweak existing practices. It could well have been that individual students who were moved into a set that better reflected their ability, or who benefitted from improved teaching in a lower set, saw improvements in their academic attainment. However, these assumed changes in the elements of this intervention didn't manifest themselves in the whole cohort changes in attainment that the trial was designed to detect.

## Limitations

The extent of missing follow-up data was so high for this efficacy trial (37% and 53% schools lost to follow-up in the Maths and English trial respectively) that it is difficult to make any strong conclusions about the effectiveness of the intervention. The completers analysis for both primary outcomes yielded a null result, but the size of the effect was shown to be very influenced by multiple imputation and sensitivity analyses. This is expected given the large proportions of missing data and therefore likely bias in the result. Furthermore, it was challenging to get enough schools for interviews during the process evaluation so, similarly to the impact results, these may also be vulnerable to bias. One aspect of the trial that remains clear, however, is the large proportion of intervention schools that ceased to deliver the Best Practice in Setting initiative.

The low follow-up rate probably has several causes: secondary schools are generally harder to recruit to trials and to retain than primary schools; dissatisfaction with the intervention (schools reported it as restrictive); the burden on schools of developer-led data collection activity alongside pupil data collection for primary outcomes; and participation over two academic years meant that the staff changes made it harder for school's continued participation in the second year.

The eligibility criteria for this trial aimed to ensure a 'pure' control group consisting of schools that set (or stream) in the relevant subject. However, one intervention group interviewee reported that her English department had been using mixed-ability groupings for their Year 7 students prior to starting the intervention. This must have represented a school recruited in error. Table 11 indicates this was an aberration and our control group can still be regarded as 'pure'.

Given the uncertainty around the effect of the intervention, and the moderate to low fidelity of some of the key implementation measures, it is difficult to validate the Theory of Change developed by the UCL Institute of Education team. However, it is clear that some schools found the implementation activities challenging, while others did not experience the full range of benefits outlined in the 'process' chain. Despite the headteachers' willingness to take part in the trial, not all staff 'bought in' to the best practice principles of setting by attainment. We recommend that these best practice principles, or the method of sharing them with teachers, should be reviewed in light of the high numbers of schools that dropped out of the intervention.

## Future research and publications

A more conclusive ITT analysis will be possible once these students sit their GCSEs and outcome data is available on the NPD for all randomised schools. We recommend this is done. Whilst this data would also allow compliance analysis, the fidelity data itself was sparse and it will be difficult to draw firm conclusions from this. Instead, we recommend taking stock of the findings from this evaluation before moving on to further work on ability grouping. In particular, more work needs to be done to make the intervention palatable to schools and in trial design to ensure much better follow-up. Specifically, the findings from the process evaluation suggest that the developers may wish to consider:

- the use of Key Stage 2 results as the basis for allocating a student to a particular attainment set: several respondents reported that the Key Stage 2 data did not always fit with later assessments that they undertook, and/or that other factors needed to be taken into consideration, such as considering students' pastoral care needs, or to ensure a certain ethnic or gender mix in each set; and
- the location, timing and format of the regional workshops: attendance declined over the life of the intervention; this was attributed to a perception amongst some interviewees that: the venues were too far away; it would have been better if the workshops had been delivered in school time, rather than as twilight sessions; the workshops should have allowed more time for discussions of a practical (rather than conceptual) nature.

The findings from the process evaluation of the Best Practice in Setting initiative and the companion evaluation of the Best Practice in Mixed Attainment pilot trial suggest there are also wider factors to consider. These include the pressures of increasing student numbers and reducing school budgets in England, the challenges of school workload, and recent interest in mastery approaches to learning, particularly in mathematics.



## References

- Ames C. (1992) 'Classrooms, goals, structures and student motivation,' *Journal of Educational Psychology*, 84, pp. 261–71.
- Ball, S. J. (1981) *Beachside Comprehensive: A Case-Study of Secondary Schooling*, Cambridge: CUP Archive.
- Boaler, J., William, D. and Brown, M. (2000) 'Students' experiences of ability grouping-disaffection, polarisation and the construction of failure,' *British Educational Research Journal*, 26, 5, pp. 631–648.
- Carpenter, J. R., Kenward, M. G. and White, I. R. (2007) 'Sensitivity analysis after multiple imputation under missing at random – a weighting approach', *Statistical Methods in Medical Research*, 16, pp. 259–275.
- Dunne, M., Humphreys, S., Sebba, J., Dyson, A., Gallannaugh, F. and Muijs, D. (2007) 'Effective Teaching and Learning for Pupils in Low Attaining Groups' (Research Report DCSF-RR011), London: DCSF [online]. Available: <http://dera.ioe.ac.uk/6622/1/DCSF-RR011.pdf> [20 October, 2017].
- Dweck, C. S. (2008) 'Can personality be changed? The role of beliefs in personality and change', in *Current Directions in Psychological Science*, 17, 6, pp. 391–394 [online]. Available: <http://journals.sagepub.com/doi/abs/10.1111/j.1467-8721.2008.00612.x> [20 October, 2017].
- Dweck, C. S. (2010) 'The perils and promises of praise', in Ryan, K. and Cooper, J. M. (eds) *Kaleidoscope, Contemporary and Classic Readings in Education* (12th edn), Belmont, CA: Wadsworth/Cengage Learning.
- Education Endowment Foundation (n.d.a) 'Consent and the Data Protection Act: Advice for Evaluators' [online]. Available: [https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF\\_guidance\\_for\\_evaluators\\_on\\_consent\\_and\\_the\\_Data\\_Protection\\_Act\\_FINAL.pdf](https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_guidance_for_evaluators_on_consent_and_the_Data_Protection_Act_FINAL.pdf) [30 November, 2017].
- Education Endowment Foundation (n.d.b) 'Setting or Streaming', in *Teaching and Learning Toolkit* [online]. Available: <https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit/setting-or-streaming> [15 November, 2017].
- Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (n.d.) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook', [online]. Available: [https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting\\_up\\_an\\_Evaluation/IPE\\_Guidance\\_Final.pdf](https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Guidance_Final.pdf) [1 May, 2018].
- Education Endowment Foundation (2015) 'Protocol for the Evaluation of Best Practice in Grouping Students' [online]. Available: [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Protocols/Round\\_6\\_-\\_Intervention\\_A\\_Best\\_Practice\\_in\\_Setting\\_Amended\\_version.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Round_6_-_Intervention_A_Best_Practice_in_Setting_Amended_version.pdf) [20 October, 2017].
- Education Endowment Foundation (2017) 'Statistical Analysis Plan for Best Practice in Setting', NFER [online]. Available: [https://educationendowmentfoundation.org.uk/public/files/Projects/Round\\_6\\_-\\_Ability\\_grouping\\_BPS\\_SAP.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Round_6_-_Ability_grouping_BPS_SAP.pdf) [30 November, 2017].
- Flores, J. (1999) 'Tracking Middle School Students for Instruction: A Study of Homogeneous and Heterogeneous Grouping' (a Thesis Presented to the Faculty of California State University, San Marcos), [online]. Available: [Education Endowment Foundation](http://csusm-</a></p>
</div>
<div data-bbox=)

dspace.calstate.edu/bitstream/handle/10211.8/461/FloresJonathan\_Spring1999.pdf?sequence=3 [20 October, 2017].

U.K. Parliament House of Commons Education Committee (2015) 'Academies and Free Schools: Report, Together with Formal Minutes Relating to the Report' (HC 258, 2014-15, incorporating HC 981, 2013-14), [online]. London: TSO. Available: <https://publications.parliament.uk/pa/cm201415/cmselect/cmeduc/258/258.pdf> [26 October, 2017].

U.K. Parliament House of Commons (2016) 'Schools: Teaching Methods: Written Question – 44947' [online]. Available: <http://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2016-09-05/4494> [26 October, 2017].

Hallam, S. and Ireson, J. (2005) 'Secondary school teachers' pedagogic practices when teaching mixed and structured ability classes', *Research Papers in Education*, 20, 1, pp. 3–24 [online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/0267152052000341318?src=recsys&journalCode=rred20> [26 October, 2017].

Hallam, S. and Ireson, J. (2007) 'Secondary school pupils' satisfaction with their ability grouping placements', *British Educational Research Journal*, 33, 1, pp. 27–45 [online]. Available: <http://eprints.ioe.ac.uk/367/> [26 October, 2017].

Husbands, C. (2014) 'Setting by ability: what's the evidence?', IOE London blog, 4 September [online]. Available: <https://ioelondonblog.wordpress.com/2014/09/04/setting-by-ability-what-is-the-evidence/#comments> [26 October, 2017].

International Association for the Evaluation of Educational Achievement (IEA) (2011) *TIMSS 2011 Student Questionnaire*, IEA: Boston.

Ireson, J., Hallam, S. and Hurley, C. (2005) 'What are the effects of ability grouping on GCSE attainment?', *British Educational Research Journal*, 31, 4, pp. 443–458 [online]. Available: <http://emilkirkegaard.dk/en/wp-content/uploads/What-Are-the-Effects-of-Ability-Grouping-on-GCSE-Attainment.pdf> [26 October, 2017].

Kutnick, P., Sebba, J., Blatchford, P., Galton, M., Thorp, J., MacIntyre, H. and Berdondini, L. (2005) 'The Effects of Pupil Grouping: Literature Review' (RR688), London: DFES [online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.127.9357&rep=rep1&type=pdf> [26 October, 2017].

Marsh, H. W. (1990) *The Self-Description Questionnaire II Manual*, Australia: University of Western Sydney: MacArthur.

Office for Standards in Education (2013) 'The Most Able Students: Are They Doing as Well as They Should in our Non-Selective Secondary Schools?' [online]. Available: <http://www.excellenceeast.org.uk/uploads/Articles/2013/Jun/TheMostAbleStudents-OfSTEDJune2013FullReport.pdf> [26 October, 2017].

Organisation for Economic Co-operation and Development (2012) *Student Questionnaire—OECD Program for International Student Assessment 2012*, Paris: OECD Publishing.

Snijders, T. A. B. (2005) 'Power and sample size in multilevel linear models', in Everitt, B. S. and Howell, D. C. (eds) *Encyclopedia of Statistics in Behavioral Science*, 3, pp. 1570–1573, Chichester (etc.): Wiley [online]. Available: <https://pdfs.semanticscholar.org/a769/1eb67c5806e154da58a74f7b1a1bc9ccb58a.pdf> [30 November, 2017].

Slavin, R. E. (1990) 'Achievement effects of ability grouping in secondary schools: A best-evidence synthesis', *Review of Educational Research*, 60, 3, pp. 471–499 [online]. Available: <http://journals.sagepub.com/doi/abs/10.3102/00346543060003471> [26 October, 2017].

Sukhnandan, L. and Lee, B. (1998) 'Streaming, Setting and Grouping by Ability: a Review of the Literature', Slough: NFER [online]. Available: <https://www.nfer.ac.uk/publications/SSG01/SSG01.pdf> [26 October, 2017].

Taylor, B., Francis, B., Archer, L., Hodgen, J., Pepper, D., Tereshchenko, A. and Travers, M. (2016) 'Factors deterring schools from mixed attainment teaching practice', *Pedagogy, Culture & Society*, 25 (3), pp. 327–345 [online]. Available: <http://www.tandfonline.com/doi/full/10.1080/14681366.2016.1256908> [26 October, 2017].

## Appendix A: EEF cost rating

Cost ratings are based on the approximate cost per student per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found [here](#). Cost ratings are awarded as follows:

Cost rating	Description
£ £ £ £ £	<i>Very low:</i> less than £80 per pupil per year.
£ £ £ £ £	<i>Low:</i> up to about £200 per pupil per year.
£ £ £ £ £	<i>Moderate:</i> up to about £700 per pupil per year.
£ £ £ £ £	<i>High:</i> up to £1,200 per pupil per year.
£ £ £ £ £	<i>Very high:</i> over £1,200 per pupil per year.

## Appendix B: Security classification of trial findings

### Mathematics

Rating	Criteria for rating			Initial score	Adjust	Final score	
	Design	Power	Attrition*				
5	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%		Adjustment for Balance [ x ]		
4	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%				
3	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%				
2	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%	2		Adjustment for threats to internal validity [ x ]	2
1	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	41-50%				
0	No comparator	MDES > 0.6	>50%				

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = 2   
Strong experimental design and appropriate analysis, combined with high attrition (only 34% of the pupils sampled for post-testing were analysed).
- **Reason for adjustment for balance** (if made): N/A
- **Reason for adjustment for threats to validity** (if made): N/A
- **Final padlock score:** initial score adjusted for balance and internal validity = 2

\*Attrition should be measured at the pupil level, even for cluster trials.

## English

Rating	Criteria for rating			Initial score	Adjust	Final score	
	<b>Design</b>	<b>Power</b>	<b>Attrition*</b>				
5	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%		Adjustment for Balance [ x ]		
4	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%				
3	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%				
2	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%			Adjustment for threats to internal validity [ x ]	
1	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	41-50%				
0	No comparator	MDES > 0.6	>50%	0			0

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = 0   
Strong experimental design and appropriate analysis, combined with significant attrition (only 40% of the pupils sampled for post-testing were analysed).
- **Reason for adjustment for balance** (if made): N/A
- **Reason for adjustment for threats to validity** (if made): N/A
- **Final padlock score:** initial score adjusted for balance and internal validity = 0

## Appendix C: Recruitment material

### BPS FAQ

#### BPS Frequently Asked Questions (on NFER website)

#### Evaluation of best practice in grouping students Setting initiative

Project code EEFK

#### Purpose

##### Why are you doing this research?

'Ability grouping' has been a long-standing topic of contention. Evidence suggests that i) students from disadvantaged backgrounds are over-represented in low sets and streams, and ii) that students in these low groups make less progress than students in higher sets/streams, and than their counterparts in mixed attainment classes. However, the research is less clear on why this is the case, and whether it can be addressed by better practice. The purpose of the study is to evaluate the impact of different, research-informed approaches to grouping students, on students' learning outcomes. It will also explore how outcomes vary between different students according to factors such as student background, prior attainment, student gender and ethnicity.

##### Who will benefit from this research?

We hope that students in schools across the country will benefit from this research as it will inform education policies and practices in England. We also hope that your school will benefit, via access to high quality CPD and research-informed practice designed to support progress and achievement for all students.

##### Who is carrying out the research?

The research is being carried out by a team from King's College London (KCL), led by Professor Becky Francis. It is funded by the Education Endowment Foundation (EEF) and is being independently evaluated by the National Foundation for Education Research (NFER). NFER are also helping to recruit schools to the study.

#### Roles

##### What is NFER's role in the study?

NFER are evaluating the project. They will randomly allocate schools to the intervention or control group. NFER will conduct telephone interviews with some teachers. NFER will administer tests developed by GL Assessment in English ('Progress Test in English') and maths ('Progress Test in Maths') to a sample of students at the end of Year 8. NFER will supply additional copies of the tests to schools on request. NFER will use the data from the tests and staff questionnaires to evaluate the

intervention and will disseminate the headline findings of the intervention. NFER are also helping to recruit schools to the study.

**What is the Education Endowment Foundation's role in the project?**

EEF are funding the project and will help to disseminate the findings.

**What is KCL's role in the project?**

KCL has developed and will be delivering the intervention to schools. They will also be undertaking their own research, although this is separate to the evaluation NFER will be undertaking.

**What does the school need to do as part of the project?**

What the school does depends on whether it is allocated to the intervention or control group.

All schools will need to provide the research teams at KCL and NFER with data about their students which will enable NFER to administer the GL Assessment progress tests to a sample of Year 8 students in summer 2017. Schools will need to facilitate students and teachers completing questionnaire surveys and may need to make students and staff available for interviews.

Schools will need to follow instructions for setting and teach students according to evidence-based best practice for teaching students in sets. Schools will need to send two English teachers and two maths teachers to each of the twilight professional development sessions. The setting instructions relate to the number of sets (no more than four set levels), allocation of students to sets using KS2 results only, random allocation of teachers to sets and moving students between sets at specific times during the intervention according to test results.

Schools in the control group continue with their usual teaching practices.

**Eligibility**

**We currently stream students – are we eligible to participate?**

If you would be willing to change your practice to setting, you would be able to participate. Participant schools are not allowed to apply streaming.

**We divide our students into two (or more) bands by ability and group within these – are we eligible to participate?**

If you would be willing to change your practice to setting, you would be able to participate. Participant schools are not allowed to apply banding.

**We set for Maths but not for English – are we eligible for 'Best practice in setting'?**

Yes. We can recruit a smaller number of schools that set in maths only. In these schools only the maths department will participate in the trial.



## **Timeline**

### **When is the deadline for committing to the project?**

The deadline for committing to the project is Friday 12th June 2015 but we would like schools to commit to the project as soon as possible.

### **When does the study start?**

Schools will be randomly allocated to the intervention and control groups in June 2015. The first professional development events for the intervention will be in July 2015. The intervention applies to students who will start in Year 7 in September 2015.

### **When does the study end?**

The study will end in July 2017, when the participating students have reached the end of Year 8. The final report will be published on the EEF website in early 2018.

### **When will we know if we have been allocated to the control or intervention group?**

Randomised allocation to intervention and control groups will take place in June 2015. You will be informed soon after.

## **Professional development**

### **When will the professional development events be?**

Professional development events will begin in July 2015. Schools allocated to the intervention group will be given a calendar of events and venues in June 2015.

### **Where will the professional development events take place?**

Professional development events will take place at regional venues. These will be confirmed in June 2015.

### **How many professional development events will there be?**

The intervention consists of four twilight professional development sessions over two years.

## **Data collection**

### **What tests will you be using to assess students' progress in English and maths?**

We will be using GL Assessment's new Progress Test in Maths and Progress Test in English. These are designed for Year 8 students and correspond to the new National Curriculum.

**What data will be collected?**

Background data about students will be collected through the National Pupil Database. To enable us to undertake the analysis we will require Year 7 pupil names, date of birth and UPNs. NFER will contact participating control and intervention schools to collect this information in September 2015.

KCL will ask schools for students' Key Stage 2 results. KCL will collect their own data about student and staff attitudes to grouping and to learning through questionnaire surveys and interviews. Attainment data for English and maths will be collected from a sample of Year 8 students in Summer 2017, using GL Assessment's Progress Test in English and Progress Test in Maths.

**How will you handle data?**

All those involved in the project treat all personal data in the strictest confidence and no individual school, student or teacher is identified in any report arising from the trial. Full ethical approval for the study has been obtained from the KCL ethics board.

**When will the report be available?**

The report will be available in early 2018.

**Where will the report be published?**

The report will be published on the EEF website and disseminated by EEF, KCL and NFER.

**Do you need to visit the school?**

We will need to visit the school only if your school is selected for student and staff interviews.

**Do the research team have current Disclosure and Barring Service checks?**

All members of the research team who may visit schools have current DBS checks.

**Who needs to give consent for participation in this study?**

The Headteacher gives consent to take part in the trial on behalf of the school. In addition, parents/carers and students themselves will be fully informed of the data we will be collecting (e.g. from questionnaires, tests and interviews) and may choose to withdraw their data from the study. They can do this by returning an opt-out consent form which we will send to schools for circulation to parents/carers. Students can withdraw their data from the study at any time up until the end of the intervention. Although we would encourage them to take part, teachers may also withdraw their data from the study and choose not to participate in research activities.

**Who has given ethical approval for this study to take place?**

The study has ethical approval from KCL.

## **Research methods**

### **What is a randomised controlled trial (RCT)?**

A randomised controlled trial is a type of study in which the people being studied are allocated randomly either to receive an intervention or to be in a control group that does not receive the intervention, so that any effect of the intervention can be assessed by comparing outcomes for the two groups.

### **Can I choose whether I am in the intervention or control group?**

No. In a randomised controlled trial, participants are allocated at random to the intervention or control group.

## **Incentives**

### **What are the incentives for participation?**

Schools in the intervention groups will receive: free, high quality professional development, access to research on best practice, practical strategies for raising attainment, evidence-based guidance on differentiation, approaches to address Progress 8, direct involvement in research, the opportunity to access free GL Assessment maths and English tests to assess students' progress and the chance to contribute to the evidence-base for raising attainment for disadvantaged students.

Schools in the control group will receive a one-off payment of £1000 at the end of the study; the opportunity to access free GL Assessment maths and English tests, the chance to contribute to the evidence-base for raising attainment for disadvantaged students, and be among the first to receive the report of the study findings in 2018.

## **The school**

### **Who from the school will need to be involved?**

Students in Year 7 (2015-16) and Year 8 (2016-17) will participate in the study, along with their English and maths teachers. Two English teachers and two maths teachers will need to attend professional development events if the school is in the intervention group. The Headteacher must give consent for the school to participate and senior leadership team support will be needed throughout the study.

### **What expenses will be paid?**

KCL will pay standard class travel expenses to and from the professional development events.

### **How can we get involved in the study?**

You can get involved by responding to your school's approach letter which will include a form to complete and return which commits you to the project. Alternatively you can contact Asma Ullah on 01753 637432 or by email ([a.ullah@nfer.ac.uk](mailto:a.ullah@nfer.ac.uk)) if you have any queries.

**NFER's recruitment invitation letter**

RPO/EEFK/41635/2

NFER No: 1

Dear Headteacher

**'Best Practice in Setting' initiative**

**What are the best ways of grouping students to raise achievement?**

**Can better approaches to grouping students benefit students from disadvantaged backgrounds?**

**Take part in research with King's College London and the National Foundation for Educational Research (NFER) to help find out.**

**We are looking for schools interested in setting practices in Key Stage 3 English and maths. By taking part in this Education Endowment Foundation-funded research project, you will get:**

**Schools in the intervention group**

- **Free, high quality professional development**
- **Access to research on best practice**
- **Practical strategies for raising attainment**
- **Evidence-based guidance on differentiation**
- **Direct engagement in research**
- **Access to free GL Assessment tests to assess your students' progress**

**Schools in the control group**

- **£1000 at the end of the study**
- **Direct engagement in research**
- **Access to free GL Assessment tests to assess your students' progress**

**Interested?**

**Please continue reading to find out more about what's involved and how to contact us.**

### **What is the study about and who is involved?**

The 'Best Practice in Grouping Students' study evaluates the impact on students' educational progress and attainment of different approaches to grouping students. It will test an intervention which trains schools in a best practice approach to setting. The study has a particular focus on low-attaining young people - often from disadvantaged backgrounds - but seeks to promote the progress and raise the attainment of all young people. The trial is funded by the Education Endowment Foundation (EEF). The research team for the trial consists of staff at King's College London (KCL), the University of Nottingham (UoN), and Queen's University Belfast (QUB). The evaluation team for the trial consists of staff at NFER. The trial will measure the performance of pupils in years 7 and 8. The evaluation should equip school staff with teaching practices designed to enhance students' learning through setting arrangements. As part of that process we are asking a number of schools to participate in the trial. Your school is one of a representative sample of schools randomly selected for this purpose and I would like to invite your help with this work. If your school is LA maintained we have notified your LA that we will be contacting you.

### **What does participation involve?**

Schools that agree to take part will be randomly selected to receive an intervention or be part of the control group. Schools in the intervention group will receive the intervention led by the expert team from KCL in **2015-16 and 2016-17**. The intervention will involve implementing the evidence-based best practice for teaching students in set groups, and twilight sessions for two English teachers and two mathematics teachers in the school with information and materials to cascade to colleagues in their departments. Schools in the control group will receive £1000 on completion of

the trial and the opportunity to access free GL Assessment maths and English tests, and be the first to receive the final research report.

Schools are able to participate provided that they currently set students for English and/or maths in years 7 and 8. Schools that are willing to change to setting practices from streaming are also eligible.

To evaluate the intervention, NFER will randomly select 30 students to take the Progress Test in English (GL Assessment) and a further 30 students to take the Progress Test in Maths (GL Assessment) from the year 8 students on the school roll in 2016/17. NFER will administer the tests to the students at the school in summer 2017. The school will need to withdraw the students from two lesson periods and provide an appropriate place for them to take the tests. GL Assessment will mark the tests and report the students' results to the school.

All students in intervention and control schools will have the chance to complete questionnaires about their experiences and engagement with English and mathematics in 2015/16 and once again in 2016/17. English and maths teachers in intervention schools will also be offered the opportunity to complete questionnaires about practices relating to setting in 2015/16.

All schools will need to provide data on all pupils in year 7 in September 2015. We will require basic pupil information such as name, UPN and date of birth. Pupils will be sampled from these lists in the second year of the initiative to sit the GL Assessment progress tests.

#### **I am interested in taking part – what should I do next?**

Please complete the enclosed reply form and fax it to 01753 790114 or send it to us in the enclosed pre-paid envelope, or by email ([groupingstudents@nfer.ac.uk](mailto:groupingstudents@nfer.ac.uk)). We would like to receive your reply **as soon as possible**. To help us to communicate efficiently with your school, and also in view of the confidential nature of this trialling, please provide a named contact from your school staff.

For more information, please see the 'Best Practice in Setting' FAQs at: [www.nfer.ac.uk/EEFK2](http://www.nfer.ac.uk/EEFK2). I look forward to hearing from you. Please do not hesitate to contact XXX on XXX or by email XXX, if you have any queries.

Yours sincerely

XXX

Research Manager

Research and Product Operations

Parent consent letter



---

**Evidence for  
Excellence in  
Education**



## INFORMATION SHEET FOR PARENTS/CARERS

### Best practice in grouping students

#### Invitation

We would like your child to participate in this research study. The study is being led by King's College London with the help of Queen's University Belfast and University of Nottingham, independently evaluated by the National Foundation for Educational Research, and funded by the Education Endowment Foundation.

Participation is voluntary. Choosing not to take part will not disadvantage your child in any way. Before you decide whether you want them to take part, it is important for you to understand why the research is being done and what their participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask the research director using the contact details below if there is anything that is not clear or if you would like more information.

#### What is the purpose of the study?

The purpose of the study is to evaluate the impact of different groupings on students' experiences and learning outcomes. Schools group pupils in different ways; for example in groups based on their prior attainment, or in mixed attainment groups. We want to find out how the grouping arrangements in your school affect pupils' experiences and attainment. The study seeks to draw conclusions about which types of student grouping best improve students' learning outcomes and experiences.

#### Why is my child taking part?

The study requires a large sample of schools so that the various types of school in England are represented in the research. This will make it possible for us to generalise from the results for the schools in the sample to all of the schools in England. All Year 7 students in your child's school and the other schools participating in the study are being invited to take part in the research.

#### Does my child have to take part?

Your child does not have to take part. You should read this information sheet and if you have any questions you should ask the director of the study, whose contact details are given below. If you are happy for your child to take part, please keep this information sheet. If you decide that you do not want your child to take part, please complete the enclosed form and ask your child to return it to their teacher. If the teacher does not receive a completed form from you, we will assume that you are happy for your child to take part.

#### What will happen to my child if they take part?

The following information will be shared between King's College London, Queen's University Belfast, National Foundation for Educational Research, Education Endowment Foundation, their data contractor FFT Education and in an anonymised form, to the UK Data Archive.

- Information provided by your child's school (including your child's name, date of birth and unique pupil number) will be linked with information about your child from the National Pupil Database (held by the Department for Education).
- National Foundation for Educational Research may administer a GL Assessment standardised test in English and/or mathematics to your child towards the end of Year 8. In this case, your child's name, date of birth and gender will be provided to GL Assessment, who will report results confidential to your child's school for educational purposes.
- King's College London will ask your child to complete a short online questionnaire about their experiences of learning English and mathematics, once in Year 7 and once in Year 8. Their teacher will give them an opportunity to complete the questionnaire during school time. It will take your child about 30 minutes to complete the questionnaire. The online questionnaire ensures that their responses are confidential.

### **Possible benefits**

The study is intended to inform education policies and practices in England – and ultimately to benefit students in schools across the country.

### **What are the possible risks of taking part?**

There are no foreseeable risks of taking part in the study.

### **Will taking part be kept confidential?**

Data from the tests, questionnaires and interviews is regarded as strictly confidential and will be held securely until the research is finished, at which point it will be deleted. Your child's participation is entirely voluntary. If you change your mind, you are free to stop their participation. All data for analysis will be anonymised. In reporting on the research findings, we will not reveal the names of any participants or your school. There will be no possibility of any individual being linked with the data.

The UK Data Protection Act 1998 will apply to all data gathered from the tests, questionnaire and interviews. This data will be held securely within the organisations listed above. No data will be accessed by anyone other than the research team or the evaluation team. It will not be possible to link any data back to any individual participating in the research.

You may withdraw your child's data from the project at any time during the project, specifically until 31 August 2017. If you ask us to withdraw your child's data at any time before then, we will remove all traces from the records.

### **How is the study being funded?**

The Education Endowment Foundation is funding this study. For further information about the organisation, visit: <http://educationendowmentfoundation.org.uk>. The study has been approved by the King's College London Research Ethics Committee.

### **What will happen to the results of the study?**

We plan to make our research findings publicly available through events such as seminars, conferences and meetings, and through publications such as reports, articles and books.

### **Who should I contact for further information?**

If you have any questions or require more information about this study, please contact the director of the study, Professor Becky Francis, by email at [Becky.Francis@kcl.ac.uk](mailto:Becky.Francis@kcl.ac.uk) or by telephone on 020 7848 3095.



If this study has harmed you in any way or if you wish to make a complaint about the conduct of the study you can contact King's College London using the details below for further advice and information:

The Chair, Social Science and Public Policy, Humanities and Law, Research Ethics Subcommittee:  
**rec@kcl.ac.uk**

**Thank you for reading this information sheet and for considering taking part in this research.**

## OPT-OUT FORM FOR PARENTS/CARERS



### Best practice in grouping students

*Please note that you need only return this form to your child's school if you do **not** want your child to participate in the research project.*

I do **not** want my child to take part in the King's College London research project:

Best practice in grouping students.

Your Name .....

Name of child .....

Signed.....(parent/guardian)

Date .....

## School Memorandum of understanding

### MEMORANDUM OF UNDERSTANDING

#### 'Best Practice in Setting' trial

This is a memorandum of understanding between King's College London, National Foundation for Educational Research and the school:

**School name:** \_\_\_\_\_ **NFER No:** \_\_\_\_\_

#### The trial

The aim of this trial is to evaluate the impact of the Best Practice in Setting intervention on student engagement and attainment in English and mathematics, particularly that of disadvantaged students. A sample of 120 schools in England will participate in the trial, which is funded by the Education Endowment Foundation (EEF). The research team for the trial consists of researchers at King's College London (KCL), University of Nottingham (UoN) and Queen's University Belfast (QUB). The evaluation team for the trial consists of staff at the National Foundation for Educational Research (NFER). The results of the evaluation will contribute to understanding which teaching practices are effective in supporting pupil progress and outcomes, and will be widely disseminated to schools in England. Ultimately the evaluation should equip school staff with teaching practices to enhance students' learning through setting arrangements and associated practices. Full ethical clearance has been obtained for the project from the KCL ethics board.

For the purposes of evaluating the Best Practice in Setting intervention, NFER will randomly allocate the school to either an intervention group or a control group. Random allocation is the best way of separating out the effect of the intervention from the effect of individual school's qualities. To ensure that this allocation is genuinely random, it is important that the school commits to the trial before it takes place. If the school is allocated to the intervention group, KCL will deliver the intervention to schools in 2015/16 and 2016/17. If the school is allocated to the control group, KCL will send the final research report to the school and make a payment of £1000 to the school on completion of the trial. The intervention itself consists of the operation of specified, evidence-based principles in the practice of set organisation and twilight sessions for two English teachers and two mathematics teachers in the school with information and materials to cascade to colleagues in their departments.

To evaluate the intervention, NFER will randomly select 30 students to take the Progress Test in English (GL Assessment) and a further 30 students to take the Progress Test in Maths (GL Assessment) from the Year 8 students on the school roll in 2016/17. NFER will provide the tests to be administered to the students at the school in summer 2017. The school will need to withdraw the students from two lesson periods and provide an appropriate place for them to take the tests. GL Assessment will mark the tests and report the students' results to the school. The school has the option of requesting additional copies of the tests for the remainder of the cohort (although these will not be externally administered or marked). NFER will match the test data with the Department for Education's National Pupil Database and share these data with the research team, EEF, EEF's data processor the Fischer Family Trust (FFT) and, once anonymised, the UK data archive. Prior to the intervention (summer 2015) and during the intervention (summer 2016), NFER may conduct a telephone interview with the Head of English and with the Head of Mathematics about their setting arrangements and associated practice.

To contribute to the evaluation of the intervention, the research team will provide online questionnaires about their experiences and engagement with English and mathematics to the cohort of students in the school soon after they start Year 7 in 2015/16 and again shortly before they finish Year 8 in 2016/17. The research team will conduct an online questionnaire survey of the school's Year 7 English and maths teachers about practices relating to setting early in the 2015/16 academic year. The research team may also conduct interviews with a small number of students and teachers involved in the intervention (if your school is involved in this element, it will only mean one visit by researchers). KCL, UoN, QUB, NFER, EEF and FFT treat all personal data in the strictest confidence and no individual

school, student or teacher is identified in any report arising from the trial. Full ethical approval for the study has been obtained from the KCL ethics board.

## **Responsibilities**

### **The research team:**

- Acts as the first point of contact with the school for the intervention and the wider trial
- Conducts random allocation of English and mathematics teachers to sets in Year 7 and 8
- Delivers four twilight sessions to two English teachers and two mathematics teachers from the school in a regional location
- Reimburses the teachers/school for standard class travel to and from the twilight sessions
- Provides the four teachers from the school with information and materials for the intervention
- Ensures all staff from the research team visiting schools have received Disclosure and Barring Service clearance
- Conducts the questionnaire surveys of teachers and students
- May conduct interviews with teachers and students involved in the intervention
- Analyses and reports on data from the questionnaire surveys and the interviews
- Analyses and reports on data from the English and mathematics tests
- Administers a £1000 payment to the school, if it is allocated to the control group, once the Year 8 students have completed the tests.

### **The evaluation team:**

- Randomises the allocation of the school to the intervention or control group
- May conduct telephone interviews with some English and mathematics teachers in the school
- Randomly selects 60 students from the Year 8 school roll and administers the English and mathematics tests to them in summer 2017
- If requested, supplies copies of the tests to the school for up to/as many as the remainder of the Year 8 cohort, at no cost
- Liaises with GL Assessment to ensure the school receives the results of the tests
- Matches the test data with the National Pupil Database for the purposes of the trial, and supplies this information to the KCL research team
- Uses data from the tests and questionnaires to evaluate the intervention
- Disseminates the findings of the evaluation.

**The school consents to random allocation to the intervention or control group and commits to participate in the trial whether allocated to the intervention or control group.**

### **The school also agrees to the following instructions:**

1. On joining the trial, schools should apply 3 or 4 set levels in English and 3 or 4 set levels mathematics.
2. Teachers will be allocated to sets according to the research team's randomisation of those teachers allocated to teach Year 7 English and mathematics or, where necessary, according to principles supplied by the researchers (and not, for example, according to expertise or length of experience). This process will be repeated as the cohort enters Year 8.

3. Students should be set at the beginning of Year 7 exclusively on the basis of their National Curriculum Key Stage 2 results, available in July.<sup>25</sup> Key Stage 2 mathematics results should be used to assign students to mathematics sets, and Key Stage 2 English results to assign students to English sets. In the case of English, schools may elect to use the Key Stage 2 externally assessed and teacher assessed elements, or the Key Stage 2 externally assessed test results only, to allocate students to sets.
4. Schools should allocate students who join Year 7 later than the beginning of the academic year to sets according to their National Curriculum Key Stage 2 results in English and mathematics if they are available; or according to their results in other tests, gathered at the earliest opportunity.
5. Decisions about re-setting students (ie moving students between sets) should be made exclusively on the basis of school assessment results and students should only be re-set at three intervals during the 2-year intervention period: in January 2016; between July and September 2016; and January 2017.
6. Schools should provide the research team with:
  - i) Data on student prior-attainment in each English and mathematics set in Year 7 during the 2015-16 academic year and in Year 8 during the 2016-17 academic year
  - ii) Information about the assessments and results used for re-setting Year 7 students in January 2015-16 and Year 8 students in 2016-17
  - iii) Completed questionnaires from a survey of all Year 7 students in autumn 2015 and all Year 8 students in summer 2017 (questionnaires provided by the research team).

And:

- iv) Complete the proforma about setting practices in summer 2015 and 2016 and supply a list of future Year 7 student names, dates of birth and Unique Pupil Numbers to NFER via a secure portal in summer 2015.
- v) Facilitate NFER's administration of English and Mathematics tests to Year 8 students in summer 2017.
- vi) If requested, facilitate a small sample of Year 7 and Year 8 teachers and students to be interviewed by the research team.
7. Two nominated Year 7 English teachers, and two nominated Year 7 mathematics teachers<sup>26</sup> should attend four regional professional development workshops on best practice in setting (provided by the project), to ensure high expectations are applied to all students, irrespective of their set level.

---

<sup>25</sup> i.e. The provisional data. As elaborated in point 5, there is an opportunity to move pupils between sets, where necessary, later in the year.

<sup>26</sup> Preferably to include the Head of Mathematics and the Head of English.

Reply Form



National Foundation for Educational Research

RPO, The Mere, Upton Park

Slough, Berkshire, SL1 2DO

Telephone 01753 637433

Fax: 01753 790111

Email address:  
groupingstudents@nfer.ac.uk

RPO/EEFK/41635/2a

**NFER No:** \_\_\_\_

**'Best Practice in Setting' initiative in key stage 3 English and maths**

	<b>Are your details correct?</b>	<b>Please amend</b>
<b>School Name</b>		
<b>Headteacher</b>		
<b>Tel. No:</b>		
<b>Fax No:</b>		
<b>Email:</b>		

Please complete the following:

I have read the attached memorandum of understanding.

My school **can/cannot** take part in this initiative of 'best practice in setting' students in key stage 3 maths  
*(please delete as necessary)*

My school **can/cannot** take part in this initiative of 'best practice in setting' students in key stage 3  
English *(please delete as necessary)*

**Headteacher/SMT signature:** .....

**Name of contact in the school:**

Mr/Mrs/Miss/Ms/Dr.....

**Contact phone number:** ..... **Contact job title:** .....

**Contact email address:** .....

Please tick below if your school currently sets and/or streams for years 7 and 8 in English and maths.

	Yr 7 maths	Yr 7 English	Yr 8 maths	Yr 8 English
<b>Do you set?</b>				
<b>Do you stream?</b>				

NFER No: 640

**Please return this completed form in the pre-paid envelope provided  
or by fax on 01753 790114 as soon as possible.**

*If you cannot help us on this occasion, we would be grateful if you could let us know the reasons why your school cannot take part. This will help us understand the factors schools consider when deciding whether or not to participate.*

.....

.....

.....

**Thank you for your help.**



## Appendix D: Randomisation syntax

title 'EEFK-Best practice in grouping students: Int A(Setting)'.  
 subtitle 'EEFK - First blocked randomisation (15/6/15)'.  
 set printback=on.

set printback=on.

GET DATA /TYPE=XLSX

  /FILE='K:\EEFK\Research and Product Operations\Randomisation\41635 - recruited schools.xlsx'

  /SHEET=name '41635 - recruited schools'

  /CELLRANGE=full

  /READNAMES=on

  /ASSUMEDSTRWIDTH=32767.

\* Check for duplicates.

freq dfeno/format=notable.

sort cases by dfeno.

match files file=\*/first=f/last=l/by dfeno.

cross f by l.

temp.

select if any(0, f, l).

list vars=dfeno description.

\* We have 3 groups of schools those taking part in just English, just maths and both.

freq TakepartinMaths TakepartinEnglish.

numeric maths.

recode TakepartinMaths ('y', 'Y'=1) ('n', 'N', ' '=0) (else=sysmis) into maths.

numeric english.

recode TakepartinEnglish ('y', 'Y'=1) ('n', 'N', ' '=0) (else=sysmis) into english.

freq maths english.

cross maths by english.

numeric subject.

```

if (maths=1 and english=0) subject=1.
if (maths=0 and english=1) subject=2.
if (maths=1 and english=1) subject=3.
value labels subject 1'Maths only' 2 'English only' 3 'Maths and English'.
freq subject.

```

```
save outfile='I:\Temp\alldata.sav'/drop=maths english.
```

- \* Aggregate the file and randomly order the strata, then alternate the
- \* allocation of the odd number to intervention or control.

```
get file='I:\Temp\alldata.sav'.
```

```

aggregate outfile = *
/break = subject
/TotalN=nu.
list subject TotalN.

```

```

set rng=mt, mtindex=8.
compute randstrata=rv.uniform(0,1).
exe.
sort cases by randstrata.
freq randstrata.
compute strataorder=$casenum.
exe.
report format=list/vars=strataorder subject(label).

```

- \* The order of the strata is maths & English followed by maths and then english.
- \* The odd schools will be allocated in the order of intervention followed by control.

- \* Randomise maths and English schools (55) - odd school to intervention.

```

get file='I:\Temp\alldata.sav'.
select if (subject=3).

```

```
set rng=mt, mtindex=12.
compute random=rv.uniform(0,1).
exe.
sort cases by random.
freq random.
compute lineno=$casenum.
exe.
numeric group.
if lineno le 28 group=1.
if lineno gt 28 group=2.
add value labels group 1 'Intervention' 2 'Control'.
freq group.
save outfile='I:\Temp\rand maths+english.sav'.
```

\* Randomise maths only schools (43) - odd school to control.

```
get file='I:\Temp\alldata.sav'.
select if (subject=1).
```

```
set rng=mt, mtindex=8.
compute random=rv.uniform(0,1).
exe.
sort cases by random.
freq random.
compute lineno=$casenum.
exe.
numeric group.
if lineno le 21 group=1.
if lineno gt 21 group=2.
add value labels group 1 'Intervention' 2 'Control'.
freq group.
save outfile='I:\Temp\rand maths.sav'.
```

\* Randomise English only schools (3) - odd school to intervention.

```
get file='I:\Temp\alldata.sav'.
```

```
select if (subject=2).
```

```
set rng=mt, mtindex=10.
```

```
compute random=rv.uniform(0,1).
```

```
exe.
```

```
sort cases by random.
```

```
freq random.
```

```
compute lineno=$casenum.
```

```
exe.
```

```
numeric group.
```

```
if lineno le 2 group=1.
```

```
if lineno gt 2 group=2.
```

```
add value labels group 1 'Intervention' 2 'Control'.
```

```
freq group.
```

```
save outfile='I:\Temp\rand english.sav'.
```

```
add files file='I:\Temp\rand maths.sav'/file='I:\Temp\rand english.sav'/file='I:\Temp\rand  
maths+english.sav'.
```

```
freq group.
```

```
sort cases by contact_id.
```

```
SAVE TRANSLATE OUTFILE='K:\EEFK\CfS\randomisation\files for the portal\Randomisation  
A - Best Practice.xls'
```

```
/TYPE=XLSX
```

```
/VERSION=8
```

```
/MAP
```

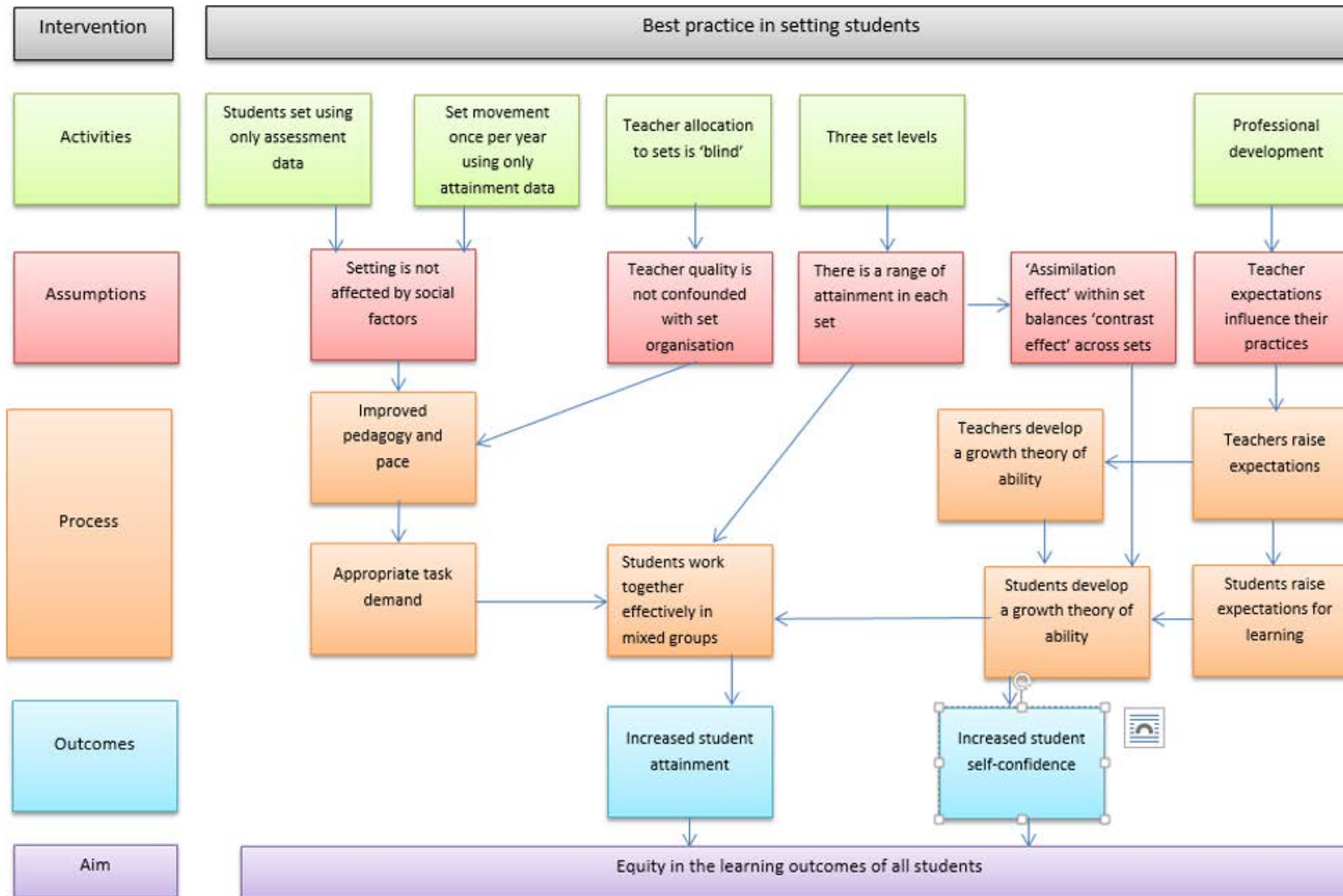
```
/REPLACE
```

```
/FIELDNAMES
```

```
/CELLS=LABELS
```

```
/DROP=random lineno.
```

## Appendix E: Theory of Change



## Appendix F: Fidelity measures

The following data was collected and/or produced by UCL Institute of Education. Key findings are reported alongside those from the telephone interviews in the process evaluation section.

### Section 1: Mathematics

#### Fidelity measure 1 – Did the expected number and type of staff attend each professional development (PD) session?\*

	PD1 (July 2015)	PD2 (September 2015)	PD3 (February 2016)	PD4 (September 2016)
Yes	53%	60%	44%	21%
No	48%	39%	56%	79%

N=61<sup>27</sup>

Percentages may not sum to 100 due to rounding.

Source: UCL Institute of Education's Register of PD attendance 2015-16.

\*It was expected that approximately four staff from each participating school would attend each PD session and that one of these would be the relevant Head of Department.

#### Fidelity measure 2 – Has some form of cascading/internal training taken place?

	PD1 (July 2015) (n)	PD2 (September 2015) (n)	PD3 (February 2016) (n)	PD4 (September 2016) (n)
Yes*	29	32	18	12
No	31	1	33	19
Data not provided	1	28	10	30

<sup>27</sup> Note one intervention school withdrew from the trial prior to the first professional development session.

N=61

Source: UCL Institute of Education's PD evaluation form 2015-16.

\*If one or more departmental members concur.

**Fidelity measure 3 – Do setting arrangements follow best practice principles?**

(n)	
Yes (3/4 sets or fewer)	29
No (5 or more sets)	9
Data not provided	23

N=61

Source: NFER's secure portal (student and set information) and NPD sample check (student KS2 results) 2015. UCL Institute of Education derived binary measures.

**Fidelity measure 4 – Were teachers allocated to classes in line with the best practice principles?**

	Year 7 (n)	Year 8 (n)
Yes, teachers randomised to classes	3	4
Partial, teachers allocated with reference to best practice principles	22	13
No, teachers not randomised or allocated with reference to best practice principles	9	0
Data not provided	27	44

N=61

Source: UCL Institute of Education's PD evaluation form and teacher proforma 2015-16.

**Fidelity measure 5 – Are 95% or more of students allocated to classes on the basis of Key Stage 2 results?**



(n)	
Yes	19
No	19
Data not provided	23

N=61

Source: NFER's secure portal (student and set information) and NPD sample check (student KS2 results) 2015. UCL Institute of Education derived binary measures.

**Fidelity measure 6 – Are 95% or more of students re-set no more than three times in two years?**

	Year 7	Year 8
	(n)	(n)
Yes	14	18
No	4	3
Data not provided	43	40

N=61

Source: NFER's secure portal (student and set information) and NPD sample check (student KS2 results) 2015. UCL Institute of Education derived binary measures.

**Fidelity measure 7 – Do teachers have high expectations for all students?**

	Baseline (November 2015)	Endpoint (June 2017)
	(n)	(n)
Yes	15	13

---

No	18	12
Data not provided	28	36

N=61

Source: UCL Institute of Education Teacher questionnaire 2015-16.

## Section 2: English

### Fidelity measure 1 – Did the expected number and type of staff attend each professional development (PD) session?\*

	PD1 (July 2015)	PD2 (September 2015)	PD3 (February 2016)	PD4 (September 2016)
Yes	47%	49%	42%	12%
No	54%	51%	58%	88%

N=43

Percentages may not sum to 100 due to rounding.

Source: UCL Institute of Education's Register of PD attendance 2015-16.

\*It was expected that approximately four staff from each participating school would attend each PD session and that one of these would be the relevant Head of Department.

### Fidelity measure 2 – Has some form of cascading/internal training taken place?

	PD1 (July 2015)	PD2 (September 2015)	PD3 (February 2016)	PD4 (September 2016)
	(n)	(n)	(n)	(n)
Yes*	16	18	5	4
No	9	1	28	14
Data not provided	18	24	10	25

N=43

Source: UCL Institute of Education's PD evaluation form 2015-16.

\*If one or more departmental members concur.

**Fidelity measure 3 – Do setting arrangements follow best practice principles?**

(n)	
Yes (3/4 sets or fewer)	17
No (5 or more sets)	5
Data not provided	21

N=43

Source: NFER's secure portal (student and set information) and NPD sample check (student KS2 results) 2015. UCL Institute of Education derived binary measures.

**Fidelity measure 4 – Were teachers allocated to classes in line with the best practice principles?**

	Year 7 (n)	Year 8 (n)
Yes, teachers randomised to classes	3	4
Partial, teachers allocated with reference to best practice principles	13	2
No, teachers not randomised or allocated with reference to best practice principles	3	0
Data not provided	24	37

N=43

Source: UCL Institute of Education's PD evaluation form and teacher proforma 2015-16.

**Fidelity measure 5 – Are 95% or more of students allocated to classes on the basis of Key Stage 2 results?**

(n)
-----

Yes	7
No	15
Data not provided	21

N=43

Source: NFER's secure portal (student and set information) and NPD sample check (student KS2 results) 2015. UCL Institute of Education derived binary measures.

**Fidelity measure 6 – Are 95% or more of students re-set no more than three times in two years?**

	Year 7 (n)	Year 8 (n)
Yes	4	8
No	2	0
Data not provided	37	35

N=43

Source: NFER's secure portal (student and set information) and NPD sample check (student KS2 results) 2015. UCL Institute of Education derived binary measures.

**Fidelity measure 7 – Do teachers have high expectations for all students?**

	Baseline (November 2015)  (n)	Endpoint (June 2017)  (n)
Yes	9	7
No	9	5
Data not provided	25	31

N=43

Source: UCL Institute of Education Teacher questionnaire 2015-16.

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.

**OGL** This information is licensed under the Open Government Licence v3.0. To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/>

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



Education  
Endowment  
Foundation

The Education Endowment Foundation  
9th Floor, Millbank Tower  
21–24 Millbank  
London  
SW1P 4QP  
[www.educationendowmentfoundation.org.uk](http://www.educationendowmentfoundation.org.uk)