



INTERVENTION	Best Practice in Mixed Attainment
DEVELOPER	UCL Institute of Education
EVALUATOR	National Foundation for Educational Research (NFER)
TRIAL REGISTRATION NUMBER	ISRCTN10304032
TRIAL STATISTICIAN	Palak Roy
TRIAL CHIEF INVESTIGATOR	Ben Styles
SAP AUTHOR	Palak Roy & Ben Styles
SAP VERSION	1
SAP VERSION DATE	15/10/2017
EEF DATE OF APPROVAL	18/10/2017
DEVELOPER DATE OF APPROVAL	18/10/2017

Protocol and SAP changes

No changes since the updated protocol was published.

Table of contents

Protocol and SAP changes	1
Table of contents	2
Introduction	3
Study design	3
Protocol amendments	4
Randomisation.....	4
Calculation of sample size	5
Follow-up	6
Outcome measures.....	6
Primary outcome	6
Secondary outcomes- pupil attitudes.....	7
Analysis	9
Primary intention-to-treat (ITT) analysis.....	9
Secondary outcome analyses.....	9
Self-confidence outcomes	10
Non-compliance with intervention	10
Subgroup analyses.....	11
Effect size calculation	11
References	12

Introduction

The Education Endowment Foundation (EEF) has commissioned UCL Institute of Education to investigate best practice in grouping students by attainment. The project is led by Professor Becky Francis and consists of two randomised controlled trials. The first trial tests an intervention which trains schools in a best practice approach to setting (BPS). The second trial is a feasibility study and pilot RCT exploring the use of mixed attainment teaching in secondary schools (BPMA).

This SAP refers to the second trial (BPMA); the SAP for BPS can also be found on the EEF website. BPMA was run as a pilot study, with the project team initially working with three secondary schools between September 2014 and July 2015. Following this developmental phase, the approach was piloted as a randomised controlled trial (RCT). NFER has been commissioned to design and manage the pilot RCT, undertake the process evaluation and to administer English and mathematics tests in summer 2017. The developers are keen to ensure that mixed attainment grouping does not result in a drop in attainment in mathematics and/or English.

Study design

Overview of the study design

This is a pilot RCT, with the project team initially working with three secondary schools between September 2014 and July 2015. Following this developmental phase the approach was piloted as an RCT, starting in September 2015 and following children through Years 7 and 8. The planned sample for the evaluation is 20 secondary schools, randomised to either receive the intervention or to be part of a control group.

NFER will administer English and mathematics tests in summer 2017 when pupils are in Year 8 to ensure that mixed attainment grouping does not result in a drop in attainment in mathematics and/or English.

Description of population including eligibility criteria

The population for this research is all state-funded English secondary schools. At recruitment for the trial, it was envisaged that most will presently employ mixed attainment grouping for one or both key stage 3 English and mathematics. However, to support the trial recruitment any school was eligible to take part regardless of their prior grouping arrangements. Since information on schools' grouping arrangements is not routinely available on school-level datasets, schools needed to complete a proforma at recruitment stage indicating the details of their grouping policy.

Sample size

The planned sample size for the evaluation was 20 Secondary schools, randomised to either receive the intervention or to be part of a control group. The sample size was larger than a conventional pilot trial as it was determined by the need to run it in enough schools to demonstrate that recruitment to a larger trial would be possible and any variability in the schools' approaches can be captured qualitatively. In addition, it is powered to detect a moderate effect size (in either direction) to inform decisions about whether to proceed and how large a future trial should be.

Description of trial arms

Intervention group: The application of the intervention involved presentation and application of the key pedagogic principles such as principles to be applied in grouping pupils and principles to be applied across groups. This was facilitated by at least six intensive workshops attended by each key teacher in the pilot study, with additional coaching/critical friend support in school. These workshops outlined and justified the principles to be applied; and to provide time, space and guidance for Year 7 mathematics and English teachers to design appropriate grouping approaches, support to pupils, and curriculum, for application of the principles in their own school context.

Control group: Schools from this group continued with their existing student grouping practices. Once they have completed the year 8 tests at the end of the trial, they will receive £1,000.

Number and timing of measurement points

The development phase took place in the academic year 2014/15 while the pilot trial spans over two academic years (2015/16 to 2016/17) where the outcome measurement will take place once at the end of academic year 2016/17.

Protocol amendments

As planned in the initial protocol, it was not possible to collect year 7 pupil data such as pupil names, DOBs and UPNs before randomisation as schools were required to administer parental consent to opt out from the data collection. Therefore, this commenced after the randomisation in academic term Autumn 2015. However, the baseline measure for the primary outcome is the pupil attainment at Key Stage 2 which took place prior to randomisation.

Another deviation from the protocol was to allocate schools to randomisation groups using an uneven ratio. This was chosen so that there are at least 10 schools in the intervention group. Please see below paragraph for details.

Randomisation

The planned sample size was 20 and the original design meant that as a result of cluster randomisation, half the schools would be randomly allocated to each group. However, it was difficult to recruit 20 schools. Only 18 schools were recruited and randomised for this trial. This meant fewer than 10 schools would have been allocated to each group as a result of school-level randomisation. The team at UCL IOE were keen to deliver the intervention to 10 schools. Therefore, it was decided that NFER would undertake random allocation unequally, where schools were allocated to intervention and control group with a 2:1 ratio.

Randomisation was carried out by a statistician at NFER using a full syntax audit trail. This was done in two blocks due to staggered school recruitment and intervention workshops running concurrently. Randomisation was conducted in June-July 2015.

In total, eighteen schools were randomised where 12 schools were assigned to the intervention group and six schools were assigned to the control group. Of these, one control school was randomised due to an administrative error (the school never intended to take part) and subsequently was removed from the trial. Four intervention schools withdrew participation and were retrospectively classed as 'not eligible'. Three of these schools approached the UCL IOE team to indicate that they were not planning to operate mixed

attainment grouping. One further school withdrew participation since they had not thought they were signing up to a trial. Since this is a pilot RCT, it was agreed that these schools will be recognised as randomised but will be classed as 'not eligible'. As they withdrew from the trial, no testing was possible in these schools so they were excluded from follow up analyses (including ITT analysis). Please see below table for details on the blocked randomisation.

Table 1: Number and proportion of schools randomised

Block	Group	n (%) at randomisation
Block 1	Intervention	10
	Control	5
Block 2	Intervention	2
	Control	1
Total	Intervention	12 (67%)
	Control	6 (33%)

Calculation of sample size

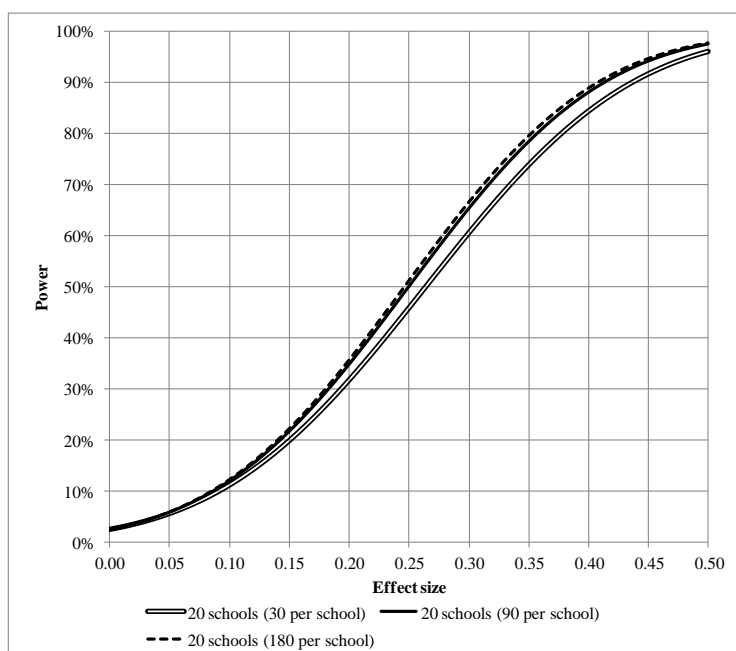
A number of within school sample sizes were considered for analysis. In order to reduce testing burden per school without sizeable impact on the power, it was decided that NFER will randomly select 60 pupils from the year 8 school roll from each of the recruited schools. Half of the pupils will sit the mathematics test and half will sit the English test¹.

Power calculations use the same assumptions as the BPS trial: intra-cluster correlation of 0.15 (lowered from 0.2 through the use of key stage 2 as a covariate); correlation between key stage 2 and year 8 test of 0.72 and average cohort size of 180. Figure 1 demonstrates how the power is very similar for a random sample of 60 (30 for each of English and maths) in each school or for sampling more. This pilot trial will therefore randomly select 30 pupils from the cohort to take an English test and 30 others to take a maths test.

¹ 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.

² At the time of writing the protocol, EEF guidance on pre-test correlation was not available. Therefore, we had used ICC and pre-test correlation based on our own research.

Figure 1 Power curves for 10 schools in each group (as per the protocol)



Follow-up

As stated earlier, pupil data could not be collected prior to randomisation. Pupil data collection started in autumn 2015. Out of eight intervention schools, six schools provided pupil data. Two intervention schools have decided not to take part in outcomes measurement and therefore will be considered complete withdrawals. All five control schools have provided pupil data.

At present, NFER is arranging the end-point test administration. The final sample with follow-up test data will be comprised of all schools that take part in this test administration.

Outcome measures

Primary outcome

While this is not a fully powered trial, testing is planned to take place in year 8 after two years of the intervention to test whether mixed attainment grouping results in a drop in attainment in mathematics and/or English³. As the intervention is aimed at pupils in years 7 and 8, testing is necessary as there is no statutory assessment in these years. Testing will take place in June-July 2017 at the end of year 8. GL Assessment's Progress in English (PTE13)⁴ and Progress in Mathematics (PTM13) tests will be used to measure the primary outcomes in English and in maths. The primary research questions of attainment will be:

1. What is the impact of mixed attainment grouping on pupils' attainment in mathematics?
2. What is the impact of mixed attainment grouping on pupils' attainment in English?

³ This will be important to establish as it will inform future decisions about whether to take the intervention to a full efficacy trial.

⁴ 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 Progressm in Mathematics (NPiM). After the development, these tests are called Progress in English and Progressm in Mathematics.

As these tests have a broad coverage of the curriculum, we will use the raw total score for each subject that covers all curriculum content. Maths total score (maximum possible score 70) will consist of fluency in facts and procedures, fluency in conceptual understanding, mathematical reasoning and problem solving. English total score (maximum possible score 66) will consist of spelling, grammar and punctuation, reading comprehension: narrative and non-narrative.

NFER will take responsibility for collecting and delivering PTE13 and PTM13 in paper form while the tests will be marked by GL assessment blind to treatment allocation. NFER will randomly select 60 pupils from the year 8 school roll from each of the recruited schools. Half of the pupils will sit the mathematics test and half will sit the English test. In order to ensure that pupils selected for the GL assessment tests include FSM-eligible pupils, this randomisation will be stratified by pupil FSM status.

Secondary outcomes- pupil attitudes

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

1. What is the impact of mixed attainment grouping on pupils' self-confidence in mathematics?
2. What is the impact of mixed attainment grouping on pupils' self-confidence in English?

These are measured by administering a pupil survey at the start of year 7 in September 2015 (baseline survey administered post randomisation) and at the end of year 8 in summer 2017 (follow-up survey). UCL Institute of Education are responsible for administration of these surveys, and response data is modelled by Queen's University Belfast. The surveys will be administered with an entire cohort in all participating schools. As participation in this trial is for both the subjects- English and maths, all the pupil surveys will be used to measure the secondary outcomes of self-confidence.

In partnership with Queen's University Belfast, UCL Institute of Education has developed pupil self-confidence measures in maths and English.

Self-confidence measures are developed using factor analysis on selected items from the baseline pupil survey data (combined dataset for both the trials, BPS and BPMA). These items were drawn from several instruments previously used (SDQII from Marsh, 1990; TIMSS questions from IEA, 2011 and PISA questions from OECD, 2012). Please see below table for the list of items included in the principal axis factor analysis.

Table 2: 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 [sic] self-concept
	I am not very good at Maths	Adapted from Marsh (1990) verbal [sic] 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 [sic] self-concept
	I get good marks in maths	Adapted from PISA self-concept in mathematics and Marsh (1990) verbal [sic] self-concept
	I learn things quickly in maths lessons	Adapted from TIMSS self-confidence in learning mathematics and Marsh (1990) verbal [sic] 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 [sic] 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

Subsequently, all the items were retained and the composite measures was created as an average of all constituent items. Self-confidence in maths had an internal reliability (Cronbach's α) of 0.88 and self-confidence in English had an internal reliability (Cronbach's α) of 0.86. Scores for these composite measures will range from 1 to 5 with higher scores reflecting higher self-confidence in the given subject.

Analysis

The trial analysis will follow EEF Analysis Policy⁵.

Primary intention-to-treat (ITT) analysis

The analysis of any pilot study should be mainly descriptive or should focus on confidence interval estimation, depending on the objectives of the study (Lancaster et al., 2004). The protocol states that intention-to-treat analysis of year 8 tests will initially analyse school means as there will not be enough schools to use a multi-level model. Whilst it is indeed the case that it is not possible to estimate between-school variance with this many schools, we have decided to retain the analysis method that might be used for a subsequent fully powered trial. This will constitute secondary analysis.

School means⁶ analysis will be the primary outcome measure of the trial. In this analysis, we will use the school averages of the PTE13 and PTM13 scores in two separate regression models. This analysis will determine whether Best Practice in Mixed attainment initiative had an overall effect on Year 8 pupils' mathematics and English attainment. For the English model, PTE13 school average will be the dependant outcome in with the following covariates:

- an indicator of whether the school is an intervention school
- school's mean prior attainment as measured by KS2 English point score (KS2_KS2READPS variable).

Similarly, model for the maths outcome will include the mean PTM13 score as a dependant variable regressed on the following covariates:

- an indicator of whether the school is an intervention school
- school's mean prior attainment as measured by the KS2 maths point score (KS2_KS2MATPS variable).

Secondary outcome analyses

The secondary analyses will include attainment and self-confidence models. These will be analysed using the multilevel models taking account of the clustering.

Attainment outcomes

Multilevel models with two levels (school and pupil) will be used for the analysis to account for the cluster randomisation. All schools with pupil data on the GL assessment English and maths outcomes will be included in this analysis irrespective to whether or not the schools implemented the intervention. There will be two separate models, one for each subject⁷. The dependent variable for the mathematics model will be the raw total score in mathematics for PTM13 with the following covariates:

- an indicator of whether the pupil is in an intervention school

⁵https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analyses_for_EEF_evaluations.pdf

⁶ School means will be calculated using the assessment data for the randomised pupils only. e.g. school's prior attainment in English will be the mean KS2 English point score for the randomised pupils only.

⁷ Note that pupils taking the English tests are different from those taking the maths tests but they will be from the same schools.

- pupil prior attainment as measured by KS2 Maths point score KS2_KS2MATPS variable.

The dependant variable for the English model will be the raw total score in English for PTE13 with the following covariates:

- an indicator of whether the pupil is in an intervention school
- pupil prior attainment as measured by KS2 English point score (KS2_KS2READPS variable).

Self-confidence outcomes

Completers analysis will be run using the secondary outcome measures as dependant variables in two separate multilevel models. The covariates for these models will be similar to the attainment models wherein pupil self-confidence measures in given subject at baseline will be one of the covariates instead of prior attainment measures.

Data manipulation will be carried out in SPSS while the multilevel models will be run in R package nlme.

Non-compliance with intervention

The developer would collect data on level of school engagement throughout the two-year delivery period using a number of pre-defined variables. They will send us data on these individual variables and we will summarise them according to pre-agreed categorisation. This categorisation will yield four measures that are listed in the following table. Measures for English are provided below as an example. Similar information would be collected for mathematics.

Table 3: Dosage variables for primary and secondary outcomes on English:

Combined measure	Variable	Level of measurement
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)
Effectiveness of student	3. Students are allocated to classes so that there is a similar range of KS2 results in each class	Binary – there is a similar range of KS2 results in each class
		0= No

Combined measure	Variable	Level of measurement
allocation		
		1= Yes
High expectations	4. Teachers have high expectations for all students	Binary
		0=No
		1=Yes
Effectiveness of differentiation practices	5. Teachers apply effective differentiation techniques in the classroom	Binary
		0=No
		1=Yes

The developer has also collected information from the control schools about their grouping practices. This data will be included in the analysis. We will undertake exploratory analysis on the fidelity measures for this pilot trial. Summary statistics will be reported but no on-treatment analysis will take place.

Subgroup analyses

Although sub-group analyses are not specified in the protocol, they will be carried out on the primary outcomes if sufficient numbers are present. . We will explore the differential effect for different pupil ability levels. This will be done by adding an interaction term to the multilevel attainment models. The intervention indicator will be interacted with pupil ability as measured by above mentioned prior attainment measures at KS2.

As per the EEF guidance, there will also be another interaction model of whether a pupil has ever received free school meals (as measured by EVERFSM_ALL variable). This will be done using a model identical to the attainment outcome model but including EVERFSM_ALL and EVERFSM_ALL interacted with the intervention indicator as covariates. Analysis shall proceed as per the original attainment outcome modelling i.e. the first model shall be identical to the attainment outcome model but with EVERFSM_ALL as a covariate.

Without doubt, above models will include a small number of pupils with FSM and hence the models will be exploratory in nature. However, we will run these models to rehearse them for a full efficacy trial.

A separate analysis of FSM only pupils will also be carried out as per the EEF analysis guidance. These models will be similar to the main models of overall effect but will only include pupils who were eligible for FSM as measured by EVERFSM_ALL variable.

Effect size calculation

The numerator for the effect size calculation will be the coefficient of the intervention group from the multilevel model. All effect sizes will be calculated using total variance from a multilevel model, without covariates, as the denominator i.e. equivalent to Hedges' g.

Confidence intervals for each effect size will be derived by multiplying the standard error of the intervention group model coefficient by 1.96. These will be converted to effect size confidence intervals using the same formula as the effect size itself.

References

IEA (2011). *TIMSS 2011 Student Questionnaire*. Boston: IEA:

Lancaster, G.A., Dodd, S. and Williamson, P.R. (2004). 'Design and analysis of pilot studies: Recommendations for good practice'. *Journal of Evaluation in Clinical Practice*, **10**, 2, 307–312 [online]. Available: <http://www.methods.lancs.ac.uk/pdfs/Lancaster%20et%20al.%202004%20JECp%20Pilot%20studies.pdf> [11 October, 2017].

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

OECD (2012). *PISA 2012 Student Questionnaire*. Paris: OECD Publishing,