Statistical Analysis Plan for Changing mindsets to improve pupil attainment

National Institute of Economic and Social Research

INTERVENTION	Changing Mindsets to improve pupil attainment
DEVELOPER	University of Portsmouth
EVALUATOR	National Institute of Economic and Social Research (NIESR)
TRIAL REGISTRATION NUMBER	ISRCTN14702744 DOI 10.1186/ISRCTN14702744
TRIAL STATISTICIAN	Dr Matthew Bursnall
TRIAL CHIEF INVESTIGATOR	Dr Matthew Bursnall
SAP AUTHOR	Dr Matthew Bursnall
SAP VERSION	V1
SAP VERSION DATE	08/02/2018

This analysis plan was written post-randomisation and prior to receipt of any outcome data, and deals only with the statistical analysis of effectiveness.

Protocol and SAP changes

Changes from current version of the protocol included in this SAP

There have been two changes:

- Pre-trial non-cognitive measures were planned to be completed prior to randomisation, as part of schools' requirements to be entered into the trial. However, to avoid delays in the implementation of the intervention, the randomisation has been done before the pre-test of cognitive measures were completed. Schools were not told the outcome of the randomisation process until they had informed the research team that the non-cognitive pre-test had been completed by pupils.
- The protocol states that the primary outcome measures would be literacy and numeracy. The Department for Education now uses the terms maths and English and has various test papers for each. As such we have revisited the choice of primary outcomes. We have changed 'numeracy' to a measure provided by DFE which combines the scores from the 3maths papers.. We have replaced 'literacy' with two measures: 'reading' and a measure provided by DFE which sums the scores for grammar and spelling. The previous changing mindsets evaluation[1] only provided

---

evidence on grammar but noted that this was a mistake and that both 'reading' and 'grammar and spelling' should have been included in hindsight. Because a key rationale for this trial was to build on the findings of the previous smaller trail, we have decided to use both 'reading' and 'grammar and spelling' as separate primary outcomes so can compare the new 'grammar and spelling' with those from previous trial and provide new evidence on 'reading'.

Introduction

The Changing Pupils' Mindsets project is testing Theory of Intelligence developed by Carol Dweck and colleagues about the theories that children hold about their intelligence, in particular whether it is a 'fixed entity' or a 'malleable' quality that can be developed. The project is an effectiveness trial of the pilot EEF project Changing Mindsets, and is implemented in primary schools with Year 6 pupils. The programme was developed by Growing Learners a team of education research psychologists, from the Department of Psychology at the University of Portsmouth, led by Dr Sherria Hoskins[2].

This trial combines two interventions from the previous trial (which showed promising results) - 1) training pupils using a structured series of learning resources and activities in the intervention developed (it was 6 weeks at that point, plus 4 weeks of local charity input but has since undergone a redeveloping and refining process) and 2) training teachers in Mindset theory and offering tips for every day practice. The aim of the current project is to train teachers to deliver the 8 weeks of intervention to increase longevity and cost-effectiveness. Teachers trained are those in Year 6 (and year 5 teachers –as the waiting control). Specifically, for schools in the intervention group, Head teachers and Year 6 teachers will be invited to attend one of approximately 10 training days hosted across the UK in September/October 2016. For schools in the wait-list control, Head teachers and Year 5 teachers were invited to attend one of 7 training days hosted at different locations across the UK at different times in Summer 2017.The intervention is 8 weeks (about 2 hours a week) long. Teachers are provided with supporting materials such as intervention manual, lessons plans, learning materials and videos.

The primary purpose of the analysis is to establish whether a Mindset intervention can increase Y6 Reading scores (as measured by Key Stage 2 Standard Attainment Tests) and self-regulation (subscale from the intervention on non-cognitive skills, captured through the Motivated Strategies for Learning Questionnaire (MSLQ). The secondary aim of the analysis is to establish whether a Mindset intervention can increase Y6 Maths scores (as measured by Key Stage 2 SATs) and self-efficacy (subscale from the MSLQ).

Study design

This is a two arm cluster randomised trial with randomisation at the school level, with schools participating in the trial randomly assigned to either the intervention group or the control group. School level randomisation was chosen over class or pupil level to minimise the chance of

---

[2] http://www.port.ac.uk/department-of-psychology/community-collaboration/growing-learners/

contamination of control by treatment. 100 primary schools were targeted, based on the power calculation in the sample size section below, to be recruited in the South East, South West, Midlands, North East and North West, with about 4,000 pupils. The intervention will be targeted at Year 6 pupils due to sit their Key Stage 2 (KS2) assessments in the academic year 2016/17. The achieved sample size is of 101 schools and 4,858 pupils.

Schools were eligible to participate in the study if:

- They had not used a systematic mindsets programme with their Year 6 cohort before.
- Schools should also be able to attend one of the training dates provided by Portsmouth.
- There was a high degree of interest from schools then the delivery team agreed to prioritise schools with a high number of Free School Meal (FSM) pupils, or schools were there was a significant gap between FSM and non-FSM pupils.
- One private school expressed an interest but was not included in the trial due to the potential impact of additional heterogeneity on the robustness of impact estimates

As a condition of entered the randomisation schools needed to provide the following:

- Head teachers consent to their school taking part in the trial (loco parentis)
- Confirmation that parent opt-out consent forms have been sent out.
- Pupil data form including UPNs for all year 6s.
- Baseline teacher questionnaire.
- Pre-test non-cognitive measures would be completed.

The design is a two arm trial, with schools participating in the trial randomly assigned to either the intervention group or the control group. All schools will pay £500 to receive the intervention. All schools will receive the full Changing Mindsets training. Schools in the intervention group will receive their training in September/October 2016, while control schools, all included in a wait-list, will receive the training two academic terms later.

The primary outcomes will be Reading and numeracy as measured by the national Key Stage 2 tests at the end of the first year.

The secondary outcome will be a measure of the impact of the intervention on non-cognitive skills, captured through the Motivated Strategies for Learning Questionnaire (Pintrich & Van De Groot, 1990). Specifically, the secondary outcome measure will focus on two subscales of the MSLQ: self-efficacy and self-regulation, and will be measured twice: pre-test non-cognitive measures would first be completed prior to randomisation; post-test using the same measures would be administrated prior to Easter 2017.

Schools were asked to provide unique pupil number (UPN). These data will allow us to request data for these pupils from the NPD, and in the future for the EEF to be able to track longitudinal data on these students.

This document has been written based on information contained in the study protocol dated 4th April 2016, in which full details of the background and design of the trial are presented.

Randomisation

An independent NIESR consultant (Dr Richard Dorsett) used STATA 13 to randomise schools 50:50 to the intervention group or to the control group after school recruitment and consent

had been received. Randomisation was within blocks defined according to tertile of school-level Key Stage 1 (KS1) performance within each of five locations: Midlands; North East; North West; South East; and South West; variable. This resulted in 15 blocks. School data were based on the Performance table data for the academic year 2014/2015 downloaded from gov.uk [3]. The final randomisation consisted of 50 schools in treatment, and 51 schools in control.

Calculation of sample size

From protocol

The sample size associated with 0.2 standard deviations in the national average for the primary outcome variable was calculated using a power calculation. When calculating the sample, the average points score was used (the 'TAPS' variable') [2].

A minimum detectable effect size of 0.20 requires the participation of 100 schools based on the following assumptions:

proportion of schools assigned to treatment 0.50
40 children per cluster [average cohort size assumption pre-randomisation]
0.05 significance level
0.8 power
0.25 intra-cluster correlation. This was a deliberately conservative assumption of around twice the size of those published by EFA[4]. A conservative assumption was chosen to reflect the fact that we would be conducting sub-group analysis for FSM.
0.5 Proportion of cluster variance predicted by KS1. This was calculated from DFE data for 14/15[2])

The minimum detectable effect size slightly decreases to 0.19 when considering the achieved sample:

proportion of schools assigned to treatment 0.50
48 children per cluster
0.05 significance level
0.8 power
0.25 intra-cluster correlation.
0.5 Proportion of cluster variance predicted by covariates.

Assuming the Free School Meal (FSM) sub-group is 15.7% of the total size of the sample (Calculated from DFE data for 14/15) and ignoring that it may be higher if recruited schools are in more disadvantaged areas), and maintaining all other assumptions (which is likely to be a conservative approach, given lower levels of within-group variation in this sub-group), this gives 8 pupils per cluster in the sub-group analysis and an estimated minimum detectable effect size for this group of 0.27 standard deviations.
The correct figures will all be reported in the final report based on 2016/17 school census data that will be requested from the DFE for the evaluation.

---

[3] "Find and compare schools in England: Download Data". https://www.compare-school-performance.service.gov.uk/download-data
[4]https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/ICC_2015.pdf

Follow-up

The delivery team asked the schools to provide with UPNs for all pupils. This has been done by administrating an opt-out consent form to all parents, explaining the current project, as well as how, and why the UPNs were collected. The delivery team recorded information on pupils whose parents decided to opt-out, therefore those pupils will not be included in the final analysis. School withdrawals will be classified as: "withdrawn post-randomisation with data received" (and included in the ITT analysis); and "withdrawal post-randomisation no data received" (and excluded from the analysis). The flow chart will summarise as much as possible numbers and reasons of withdraws.

To date, 6 schools dropout with no data available for analysis because of school opt-out. Of them, 2 are treated schools (86 pupils) and 4 are control schools (207 pupils). Moreover 11 parents provided the opt-out form (6 treatment, 5 control). The current number of pupils is 4,554. Details are provided in the CONSORT chart (Figure 1).

Figure 1: CONSORT diagram of participants through the study
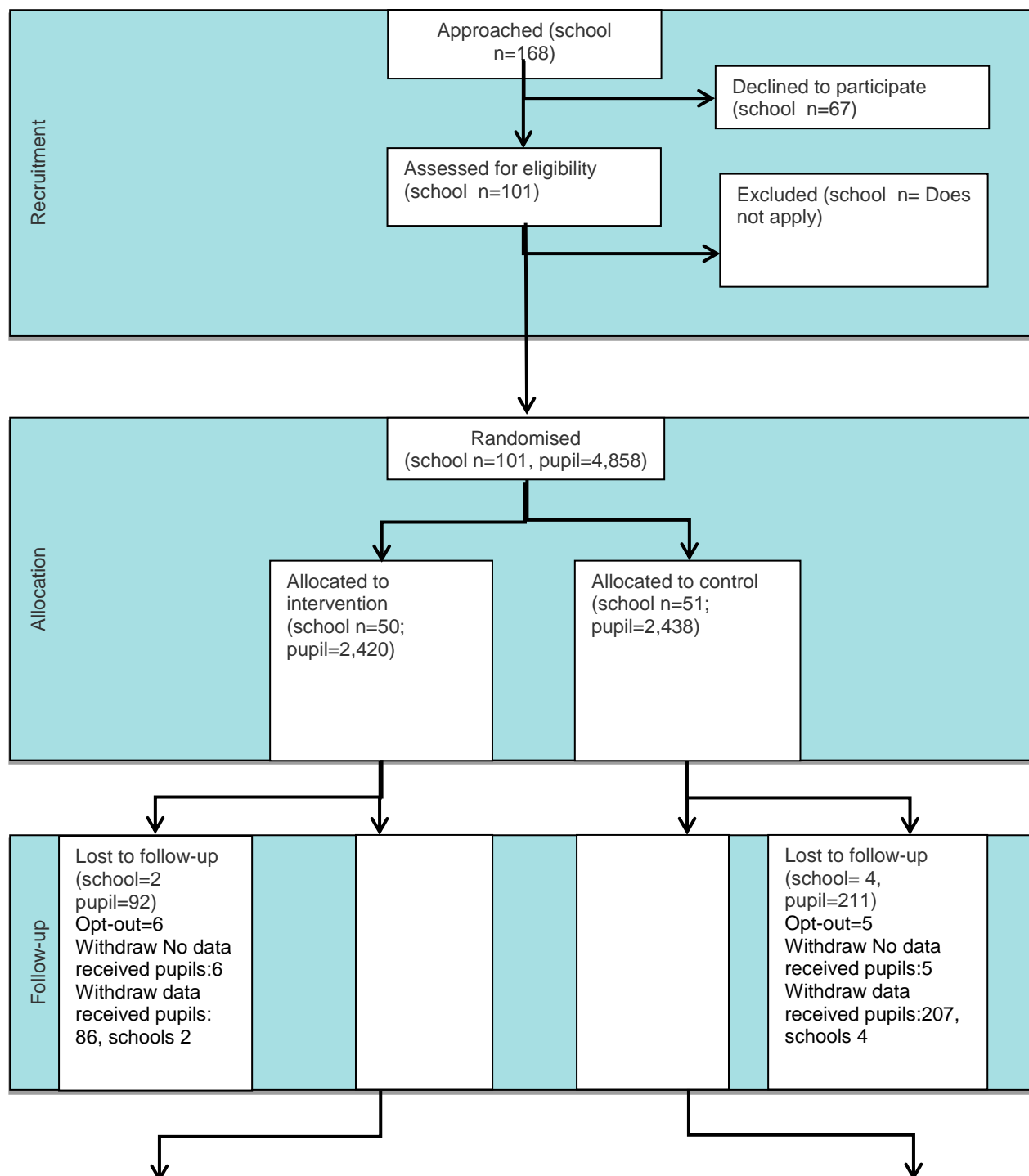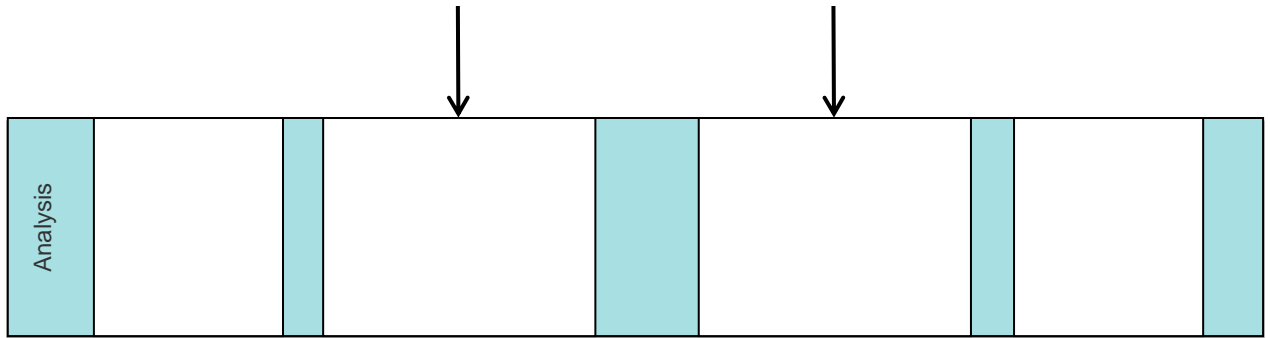
Outcome measures

Primary outcome

The primary outcomes will be reading and numeracy as measured by the national Key Stage 2 tests at the end of the first year. The primary outcome variables will be based on the National Pupil Dataset. KS2 are mandatory, national tests, and will be used as the outcome measures in this trial. KS1 scores (from tests that will have been sat at the end of the academic year 2012/13) will be used as measures of prior attainment in these subjects.

Specifically KS1 control variables for the English outcome will be KS1_READPOINTS, and for Maths will be KS1_MATPOINTS

As schools are required to conduct these tests, missing data levels are expected to be relatively low and related to absence or missing papers, although there is a risk of higher missing data recently because some parents have not been allowing pupils to sit the KS2 tests. This should not be a problem because of the conservative assumption for Intra Class Correlation coefficient used in the sample size calculation.

Specifically the main analysis will be based on the following:

- Ks2_readscore for Reading
- Ks2_matscore for Maths

We will request data from the National Pupil Database (NPD) for all pupils in the Year 6 from participating schools (treatment and control), and in particular will use the variables detailed in Table 1. The primary outcome measures will be ks2_readscore, ks2_gpsscore and ks2_matscore .

Table 1 Description of main variables based on NPD

NPD Alias	Description	Values	Use in Analysis
KS2_READSCORE	Scaled score in reading	0-999	Primary
KS2_MATSCORE	Scaled score in maths	0-999	Primary
KS2_GPSSCORE	Scaled score in Grammar, Punctuation and Spelling	0-999	Primary
KS2_PRREADLEV	Prior KS1 reading level	A, D, W, 1, 2C, 2B, 2A, 3, 4	Prior attainment control variable
KS2_PRWRITLEV	Prior KS1 writing level	A, D, W, 1, 2C, 2B, 2A, 3, 4	Prior attainment control variable
KS2_PRMATLEV	Prior KS1 maths level	A, D, W, 1, 2C, 2B, 2A, 3, 4	Prior attainment control variable
KS2_FSM6_P	Is pupil known to have been eligible for FSM for any period in the last 6 years?	0 = No 1 = Yes	Control variable

Notes:
Based on https://www.gov.uk/government/publications/national-pupil-database-user-guide-and-supporting-information
All variables are Tier 4 except KS2_FSM6 which is Tier 2

Secondary outcomes

The secondary outcome will be a measure of the impact of the intervention on non-cognitive skills, captured through the Motivated Strategies for Learning Questionnaire (MSLQ[5]). Specifically, the secondary outcome measure will focus on four subscales of the MSLQ: self-efficacy; intrinsic value, test anxiety and self-regulation. The standard MSQL questions were adapted by Portsmouth University to make them relevant to the age and location (i.e. UK) of participants based on- Moote et al[6]. Schools administered the pre-trial questionnaires to pupils when they were in Year 5 (March – July 2016), The post-trial questionnaires were completed after KS2 SATS (May – July 2017). After processing the data will be provided to NIESR for the impact evaluation.

All items on the adapted version of the MSLQ are scored from 1-7, with 1 indicating 'Strongly disagree' and 7 indicating 'Strongly agree'. (NB 3 items on the self-regulation scale need reversing before totalling the sub-scale score and the total score).

Sub-scales (Intrinsic value, self-efficacy, test anxiety and self-regulation) are then calculated as a mean score.

Analysis

The two types of schools included in the trial are:
    a) intervention schools that deliver Changing Mindsets

---

[5] http://stelar.edc.org/instruments/motivated-strategies-learning-questionnaire-mslq

[6] Moote, J, et al. 2014; When Students Take Control: Investigating the Impact of the CREST Inquiry-Based Learning Program on Self-Regulated Processes and Related Motivations in Young Science Students.
https://www.researchgate.net/publication/236606231_When_Students_Take_Control_Investigating_the_Impact_of_the_CREST_Inquiry-Based_Learning_Program_on_Self-Regulated_Processes_and_Related_Motivations_in_Young_Science_Students

b) control schools that will receive the intervention one year later

The estimated impact will be based on the difference between a) and b), regardless of drop out by intervention schools. This is in order to estimate the "intention to treat" (ITT) effect.

Analysis will be conducted in Stata v13 using the principles of intention to treat, where data are available, including all schools and pupils in the groups to which they were randomised and only including prior attainment and the allocation dummy as covariates in line with the most recent EEF guidance (https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_for_EEF_evaluations.pdf, accessed on 01/06/2017). A CONSORT diagram will be produced to show the flow of schools and pupils through the trial.

Statistical significance will be assessed using two-sided tests at the 5% level. Estimates of effect with 95% confidence intervals (CIs) and p-values will be provided.

Primary intention-to-treat (ITT) analysis

The primary objective of this study is to investigate the effectiveness of the intervention on the literacy and numerical skills of the participating pupils. Unadjusted raw and scaled outcome scores will be summarised by trial arm. The difference in reading, 'grammar and spelling', and maths attainments between identified pupils in the intervention group and those in the control group will be compared using a mixed model with school effect as a random variable, KS1 scores and indicator dummies for the stratification groups as fixed effects in STATA using restricted maximum likelihood estimation. Estimates will be done separately for each outcome.

In addition, models including only the treatment assignment and KS1 scores will be computed as required by EEF. Any differences observed between both approaches will be explained.

Grand and group means will also be reported as exploratory data analysis and the impact of KS1 scores will be reported as an effect size.

Model equation:

$$y = X\beta + Z\mu + \epsilon$$

Where:

$y = \beta X + + Z\mu + \varepsilon$

y = vector of outcome scores

X = covariate matrix [KS1 scores in 'simplest' model and this plus dummies for stratification groups in the 'precise' model)

Z = design matrix identifying which school or cluster an individual attended.

μ = vector of school random effects

β, = fixed effect parameters

ε_{ij} = residual error term for j-th member of cluster (school) i

with the covariance structure given by $\sum$, where:

$$\Sigma = (\sigma_a^2 + \sigma_e^2)\begin{bmatrix} I & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & I \end{bmatrix}$$

Where σ_a^2 is a measure of school level variation; σ_e^2 is a measure of student level variation and *I* is given by:

$$I = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{bmatrix}$$

And ρ is the intra-school correlation coefficient:

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

The fixed effect parameters and variance components will then be estimated by restricted maximum likelihood estimation using the STATA command:

mixed KS2 KS1 TREAT i.[Tertile of school average KS1] i.Region || SCHOOL: reml - for the precise model and:

mixed KS2 KS1 TREAT || SCHOOL: reml - for the simplest model

Interim analyses

No interim analyses will be undertaken.

Imbalance at baseline for analysed groups

School and pupil characteristics and measures of prior attainment will be summarised descriptively by randomised group both as randomised and as analysed in the primary analysis (for identified pupils). No formal statistical comparisons will be undertaken (Senn, 1994). Continuous measures will be reported as a mean, standard deviation (SD), minimum and maximum, while categorical data will be reported as a count and percentage.

Missing data

We will describe and summarise the extent of missing data in the primary and secondary outcomes, and their respective control variables and reasons for missing data will be described. This is likely to be a bigger issue with the secondary outcome variable. We will trigger a full multiple imputation strategy if more than 5% of data in the simplest or precise model is missing. We will also trigger imputation if more than 10% of data for a single variable or a single school is missing. The below approach will be followed separately for each instance of model and outcome for which the threshold is exceeded.

The first step will be to assess whether the missing data is missing at random (MAR). We will use the standard approach where we create an indicator variable for each variable in the impact model specifying whether the data is missing or not and use logistic regression to test whether the missing status can be predicted from the following variables: all variables in the precise model plus school average KS1 (continuous variable as opposed to tertiles), gender, ethnicity and eligibility for FSM. Where predictability is confirmed we will proceed with MI. Where the missing-ness cannot be predicted we will assume the data is either 'Missing Completely at Random' (MCAR) or 'Missing Not at Random' (MNAR). In the first case we are

unable to observe data related to randomness and MI is not feasible. In the second case the only approach would be to adopt a structural modelling approach which we would not adopt because this would deviate from the principles of transparent reporting as findings would be assumption rather than data driven.

For the models which meet the thresholds above and the MAR assumption holds we will use all variables in the precise model plus those mentioned above and adopt an MI strategy using a fully conditional specification, implemented using STATA MI to create 20 imputed data sets. We will re-estimate the treatment effect using each dataset and take the average and estimate standard error using Rubin combination rules.

We will base confirmation of the effectiveness of the treatment on complete data points only but assess the sensitivity of the estimate to missing-ness using the imputed estimates.

If the complete data only model confirms effectiveness but the imputed estimate does not we must assume that the missing data is missing not as a random to such an extent as to invalidate our conclusion of effectiveness.

Non-compliance with intervention

The original intention was for The University of Portsmouth to collect information on the amount of tuition received by pupils (number of sessions and ratio of tutor to pupils in their group). The University of Portsmouth asked Schools to provide data on the number of pupil absent from each lesson and the extent to which they adapted the material to fit their own style. Response rates were lower than expected and decreased over time, from 49 schools in session 1 to 9 in session 8. As such it will not be possible to conduct the initially planned analysis into the effect of non-compliance due to concerns of data quality.

Secondary outcome analyses

The secondary analysis will be based on the four subscales of the MSLQ. This is an important part of the Theory of change underpinning the intervention: increasing pupil's belief in their ability to improve should translate into a better learning mind set (possibly harder work) and in turn better test results. If the impact of the treatment in terms of KS2 scores is found to be significant and positive but the analysis of MSQL suggests there has been no change in attitudes towards learnings we should question whether the improved test scores are a result of the intervention.

The questions used, colour coded by which of the 4 sub-scales they belong to, are provided in annex 1. Using the numerical value associated with the scale (1 to 7) the average for each sub-scale for each pupil will be calculated for their pre-intervention and post-intervention tests. The level of association between individual question and the composite score will be reported using Alpha reliability scores.

To assess the changes on the secondary variables we will use the same modelling approach as above, but because the questions were asked pre and post intervention and the scores are non-cognitive and not necessarily related to academic ability, we will use the post-intervention non-cognitive score for each sub-scale as outcome variables and control by the pre-intervention non-cognitive test rather than KS1 score. We believe this approach is justified because there is no evidence that non-cognitive skills correlate well with KS1 performance and the EEF guidance states that "pre-tests should only be conducted if they are needed to evidence the causal model for an intervention" – which is the intention of the non-cognitive test analysis. As with the analysis of the primary outcome variable we will use the simplest

and precise specification and compare results.

Additional analyses

We do not plan any additional analysis.

Subgroup analyses

A separate analysis of the FSM subsample will be carried out following the analytical model described before. An interaction effect model (i.e. by including FSM as a main effect and an interaction with prior ability) will also be included but the subgroup analysis will be reported as the headline as we believe that is easier for the teaching community to interpret.

Effect size calculation

Effect sizes will be calculated based on the adjusted mean difference between the intervention and control group (controlling for prior attainment) and the variance components produced by STATA using the multi-level model described above. The effect size and 95% confidence interval will be calculated using equations (19) and (20) given in Hedges (2007) for cluster randomised designed analysed via multilevel models and allowing for unequal cluster sizes.

Report tables

All results will be presented using the EEF pre-specified templates. We do not plan any deviation from the pre-specified tables.

Annex 1 – Amended MSLQ questions

Scoring of the 'School Questionnaire'

All items on the adapted version of the MSLQ are scored from 1-7, with 1 indicating 'Strongly disagree' and 7 indicating 'Strongly agree'. (NB 3 items on the self-regulation scale need reversing before totalling the sub-scale score and the total score).

Sub-scales (Intrinsic value, self-efficacy, test anxiety and self-regulation) are then calculated as a mean score.

Intrinsic value	Min / Max possible score
	1 / 7 (All)
1) I prefer class work that is challenging so I can learn new things.	
4) It is important for me to learn what is being taught in class.	
5) I like what I am learning in class.	
7) I think I will be able to use what I learn in one class in other classes.	
10) I often choose class and homework activities I will learn something from even if they require more work.	
14) Even when I do poorly on a test I try to learn from my mistakes.	
15) I think that what I am learning in class is useful for me to know.	
17) I think what we are learning in class is interesting.	
21) Understanding school subjects is important to me.	

Self-efficacy	Min / Max possible score
	1 / 7 (All)
2) Compared with other pupils in class, I expect to do well.	
6) I'm certain I can understand the ideas taught in class.	
8) I expect to do very well in class.	
9) Compared with others in class, I think I am a good pupil.	
11) I am sure I can do an excellent job on the problems and tasks given in class.	
13) I think I will receive a good mark in class.	
16) My learning skills are excellent compared with others pupils in class.	
18) Compared with other pupils in class I think I know a great deal about the subject.	
19) I know that I will be able to learn the information in class.	

Test anxiety	Min / Max possible score
	1 / 7 (All)
3) I am so nervous during a test that I cannot remember facts I have learned.	
12) I have an uneasy, upset feeling when I take a test.	
20) I worry a great deal about tests.	
22) When I take a test I think about how poorly I am doing.	

Self-regulation	Min / Max possible score
	1 / 7 (All)
23) I ask myself questions to make sure I know the information I have been learning.	
24*) When work is hard I either give up or learn only the easy parts.	
25) I do extra work and practice exercises even when I do not have to.	
26) Even when lessons are dull and uninteresting, I keep working until I finish.	
27) Before I begin school work, I think about the things I will need to do to learn.	
28*) I often find that I have been reading but I do not know what it is all about.*	
29*) I find that when the teacher is talking I think of other things and do not really listen to what is being said.*	
30) When I am reading, I stop once in a while and go over what I have read.	
31) I work hard to get a good mark even when I don't like a subject.	

* Reversing required – the raw scores on these 3 items must be reversed before totaling the sub-scale or total score.