Statistical Analysis Plan for Lesson Study

London School of Economics

For some evaluation teams, the trial manager and statistician may be the same person and for others not. In all cases, the SAP should be written for a statistician or analyst to be able to carry out the analysis without prior knowledge of the trial. This is important in order to avoid bias. Describing the analyses in sufficient detail for someone else to carry it out with certainty avoids conscious or sub-conscious decisions being made on the basis of results seen. The SAP, if written sufficiently early, also provides continuity should key members of the evaluation team leave their institution during the course of the trial.

Depending on the level of detail within the trial protocol, some sections of the SAP can be cut and pasted from it. Others will require further detail. The SAP should be written at least three months before the analysis is conducted and will be reviewed by one of a panel of EEF SAP reviewers. For new EEF projects, a SAP will be appended to the protocol at the beginning of the trial and this will be updated three months before the analysis. This template should be used in conjunction with the EEF Analysis Guidelines and EEF Report Template.

INTERVENTION	**Lesson Study**
DEVELOPER	Lesson Study
EVALUATOR	Richard Murphy
TRIAL REGISTRATION NUMBER	
TRIAL STATISTICIAN	Richard Murphy
TRIAL CHIEF INVESTIGATOR	Richard Murphy
SAP AUTHOR	Richard Murphy
SAP VERSION	
SAP VERSION DATE	06/03/2017
EEF DATE OF APPROVAL	06/03/2017
DEVELOPER DATE OF APPROVAL	06/03/2017

Introduction

It is widely accepted that teachers are the most important factor in school effectiveness. However, there is little robust quantitative evidence of the effectiveness of programmes that improve teacher quality. Of reliable research carried out on teacher development, knowledge based training has typically been found to be ineffective, while programmes that involve teacher observations that are embedded into the school alongside effective feedback have been found to have effects (Taylor and Taylor, 2013).

Lesson Study is a programme incorporating teacher observations and feedback. It is a professional development programme with a long history of use in Japan and is being increasingly used in the UK. Teachers work in small groups to plan lessons that address a shared teaching and learning goal. They then observe each other's lessons, focusing on pupil learning rather than the teacher. They then discuss the lesson, refine the lesson plans and repeat the process. In the planning, two 'case pupils' are selected that are typical of a group of pupils in the class, so that the impact of the lessons on pupils with identified barriers to learning can be monitored. After the lesson has taken place, it is discussed, analysed and evaluated by the group of teachers, leading to revisions being made in content, structure and delivery. This process includes feedback from the case pupils in the form of a teacher/pupil interview. The knowledge generated by the whole process is then shared more widely with other colleagues.

An evaluation of the UK's National Strategies' Leading Teachers Programme, which involved Lesson Study, showed that those schools using this approach (among others) out-performed a comparison group in both English and Mathematics. Lesson Study also shares many of the key characteristics of effective CPD that were identified in a systematic review produced by the Evidence for Policy and Practice Information and Co-ordinating Centre. Research investigating the National Strategies' Leading Teachers Programme (Hadfield, Jopling and Emira, 2011), of which Lesson Study was a key element, revealed positive impacts on pupil outcomes. However, this study did not take account of potential differences in the treatment and control groups used in the study, nor was it able to follow up the long-term impacts of the programme.

The trial of Lesson Study incorporated Talk for Learning, an approach which aims to improve the quality of classroom talk in order to increase pupils' engagement, learning and attainment. Participating schools were asked to deliver Lesson Study within literacy and maths lessons using Talk for Maths and Talk for Literacy with teachers trained in specific approaches for both. The trial was therefore a test of the effectiveness of this method as well as of the Lesson Study approach.

Study design

Population and eligibility criteria

The target population for this study are state primary schools in England with above average Free School Meal Eligibility (FSME) shares and two or less classes per cohort. The Lesson Study developers were asked to approach such primary schools in three regions to recruit into the study. If insufficient schools were interested in the evaluation, then the eligibility criteria with regards to FSME share, and classes per year could be relaxed. The regions

were the South West, East Midlands and North West.[1] A regional approach was chosen to minimise travel time for trainers and participants when training sessions occurred. Within these schools the target teachers were those teaching in academic years 4 and 5

Sample size

The aim of the recruitment was to eventually have 160 schools participate in the study. This total was determined by baseline power calculations. Randomisation would be done at the school level, such that there would be 80 treated school and 80 control schools (see 'Randomisation'). Teachers of the target year groups in treated schools would be trained in the Lesson Study programme. The program runs over two years (2013/14 and 2014/15), with the same schools being in the treated or control groups throughout the study.

The developers stated that we could reasonably expect an effect size in the range of 0.40. We used a more conservative estimate of the impact for the power calculations given the developer estimates we based on small-scale developer-administered studies, with a shorter follow up period, and outcomes measures that were more closely aligned to the intervention. Moreover, we cautiously assumed that that there would only be 30 students in each school. With these features the trial would be of sufficient size to detect a standardised effect size of 0.10 with 80% power.

Lesson Study was able to recruit 182 schools. As there was only funding for 160, from this number we randomly selected schools to treatment and control status (details on randomisation below). When we approached the assigned treated schools at this stage 16 no longer wanted to continue with the trial. Therefore we were able to use these freed up treatment slots on the remaining schools. Critically, both of the originally assigned pair (control and refused treatment) will remain as part of the analysis on the basis of their original assignment. For each freed up slot a random school, from as yet unassigned pair, was selected into the treatment group. This resulted in a total of 89 schools assigned to treatment, 92 to control and one was left unassigned. The unbalance in these numbers is due to some schools existing as part of a school-cluster and therefore they were randomly selected into treatment or control on a cluster basis (full details can be found in the randomisation section below).

Description of trial design

In order to have sufficient schools to meet the power calculations the trial was designed to maximise sample size by keeping costs to a minimum. This was achieved by relying on Key Stage 2 (KS2) test score data from the treatment and control schools. Once a school was assigned as a control school they received a letter stating that as a result of the randomisation process they had not been selected, and we had no further contact with them for the duration of the experiment. We are still able to collect outcome data on the students from the control schools by using the national pupil database.

Schools that were selected into the treatment group were asked to select three teachers to be part of the Lesson Study trial. These teachers were all to come from years 4 and 5. Where that was not possible, because there was only one class per cohort for example,

[1] The Local Authorities involved from each region were; *North West*- Liverpool, St. Helens, Sefton, Wirral, Manchester, Oldham, Rochdale, Stockport, Tameside, Lancashire, Cheshire West and Chester ; *East* – Luton, Central Bedfordshire, Bracknell Forest, Cambridgeshire, Peterborough; *South West*: Devon, Plymouth

schools were free to involve teachers from Years 3 or 6. This means that is possible for students to be taught by Lesson Study teachers from Year 3 through to Year 6.

Table 1 shows the targeted cohorts in academic Years 4 and 5 during the calendar school years 2013-14 and 2014-15. The outcome is measured at the end of Year 6. C1 and C2 sat their respective KS2 exams in the summer of 2015-16 and 2015-16 respectively. C1 will only have one year of treatment, C2 will likely have two years of treatment, but both will have a year without treatment before their KS2 examinations. However, for reasons explained above it will also be the case that some students will be treated in schools during Year 6. These students could have taken their KS2 examinations in 2013-14 as part of C0, but also later cohorts in 2014-15, or 2015-16 may have experienced teaching Year 6. Moreover there are likely to be students taught by Lesson Study Teachers, for whom we would not have observed their KS2 outcomes by the summer of 2015-16.

Table 1

	2013-14	2014-15	2015-16	2016-17
Year 3	C3			
Year 4	C2	C3		
Year 5	C1	C2	C3	
Year 6 (KS2 exam)	C0	C1	C2	C3

Therefore the outcome measures of primary interest are for cohorts C1 and C2 in both the treatment and control schools. Results from previous cohorts will also be collected to implement a RCT-Difference-in-Differences approach described below. We will also collect these students KS1 test scores from the NPD to use as baseline measures of achievement. These students all sat their KS1 exams before the intervention took place, so these test outcomes can serve as baseline even in the presence of across-year spillovers of the treatment within schools.

Protocol changes

Due to recruitment and time pressure in the recruitment phase, some schools that were recruited do not meet the above-average FSME, or two class size criteria. Moreover, six of the schools that were recruited are part of a school cluster and resisted the idea of randomisation within a cluster, and therefore were treated as one school for the purposes of randomisation and analysis. For the purpose of clarity they are counted as separate schools in the consort diagram. Specifically there were three clusters of two schools pairs. Each pair was either allocated to treatment or control. The primary and secondary outcomes have been revised (from including combined maths and English and separate maths and English scores for each cohort) to reduce the number of primary outcomes from six down to two (just combined scores for each cohort). This simplifies the interpretation of the outcomes.

Randomisation

Richard Murphy performed a stratified randomisation of schools by Local Authority with the aim of balancing the randomisation at LA level e.g. the paring of schools for randomisation was conducted within each Local Authority. This was to ensure there were equal numbers of treated and control schools within each region and that they would be balanced in terms of unobservable regional characteristics.

We first generated a single index intended to match similar schools, among the volunteer schools, on the basis of multiple characteristics. This composed of students' average KS2 outcomes in English and Maths and proportion of FSME students who took their KS2 in the summer of 2011. To create a single index mean values for each school in 2011 were combined using Principal Component Analysis (PCA) to form a single index. There were eight school that did not exist in their current form in 2011 (five due to not having a year 6, and four being new as a result of an amalgamation), and so had no KS2 or FSME information to create an index out of. Before the first randomisation, to avoid the potential complicating issue of school-clusters, all clustered schools (6, 3 pairs) were removed from the first randomisation pool, along with those with no index, 15 schools in all.

Schools were then sorted according to their LA and then by their index score, ensuring that similar schools were ordered next to each other. Pairs were then formed by taking the first two schools within a region, and then each subsequent pair. All remaining un-paired schools from LA with an uneven number of schools were then put into another pool and ordered according to the index. Again schools were paired according to their next school. One school was left un-matched.

A random seed was set and then each school was then given a random number from a uniform distribution. The school with the higher random number within each pair was then chosen as the treatment school. 83 schools were assigned to treatment and 84 were assigned to control. The one unmatched school was assigned to a control status. Having the unmatched school forming part of the control group was part of the research design. As any un-matched schools will not contribute to the main analysis as pair fixed effect would absorb the average test scores of students enrolled at this school. Therefore it was decided to maximise the limited budget by not carrying out the treatment in this school.

These 83 schools were invited to take part in the treatment, of which 16 no longer wanted to take part. This freed up space for all the remaining schools not included in the first randomisation to be used in a second randomisation. The second randomisation used a similar method to before, pairing schools and then selecting the school with the higher random number within a pair to be the treated group. This pairing occurred within groups of reason for exclusion from the first randomisation to ensure similarities e.g. cluster schools, amalgamations and new Year 6. From this randomisation six were assigned to treated and nine to control. This imbalance is due to an odd number of schools within the cluster grouping and the new Year six grouping, with the unmatched schools set to be a control. To be specific there were three school clusters; one cluster would be unmatched (two schools). In addition to this there were an uneven number of schools that had a new Year 6, this left another school unmatched. This left a total of three unmatched schools from the second randomisation.

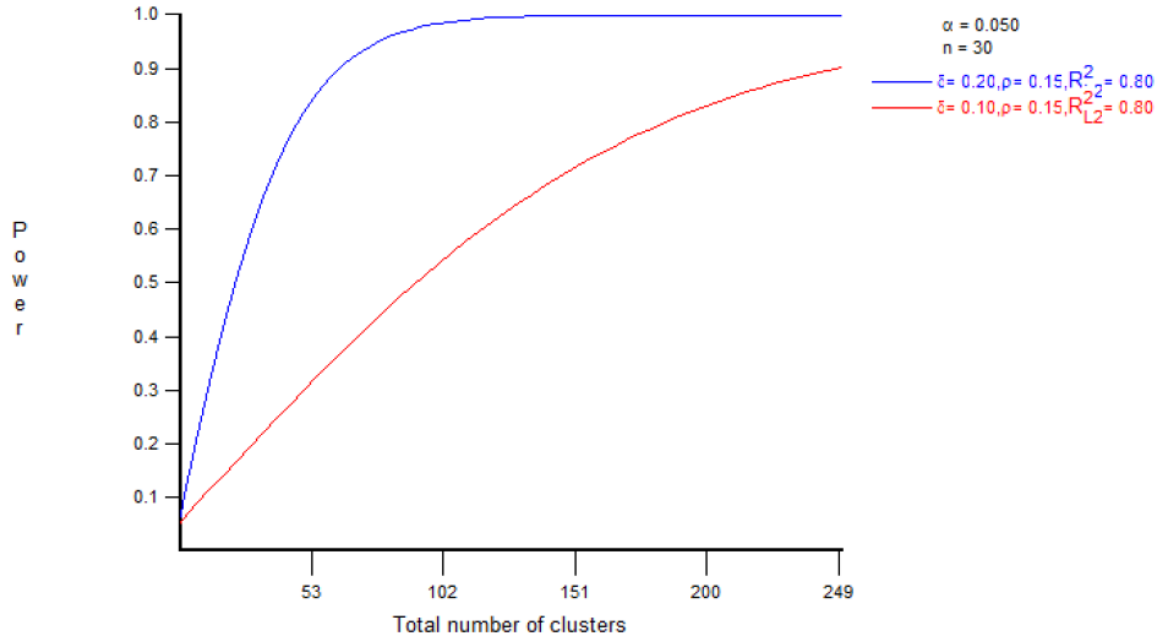
Therefore in total there are 89 school allocated to treatment, and 93 schools allocated to receive no treatment four of which are unpaired and therefore will not contribute towards the estimation of the effect.

Calculation of sample size

Assuming an intracluster correlation coefficient of $\boldsymbol{\rho = 0.15}$, covariates including prior test scores explaining 80% of the variation in test scores, a class size of 30 and requiring a 5% significance level , power analysis revealed that a large number of clusters (in our case schools) would be needed for having a good chance to detect effect sizes of 0.1-0.2 standard deviations. This was using the Optimal Design Software (Raudenbush et al., 2011). As a result, the Lesson Study trial was restricted to two arms (treatment and control) and all efforts were undertaken to include as many schools as possible, subject to budgetary constraints. The resulting sample of 182 schools (not controlling for dropouts) lies in a power region where the detection of true effect sizes of 0.2 of a s.d. in the outcome is very likely. However, an effect of this size would be considered large in the literature and the study only detects true effect sizes of 0.1 s.d. in the outcome with a probability of about 80 percent (79.4 percent with 182).

Because we could not identify students to teachers the ideal was to have only one class per cohort. However in reality many schools had two classes per year group. We used a cohort size of 30 to be conservative. Re-calculating the power calculations with 60 students within a school instead of 30, with the same set of assumptions means we reach a power of 0.8 at 144 schools, instead of 184 with an effect size of 0.1.

Figure 1

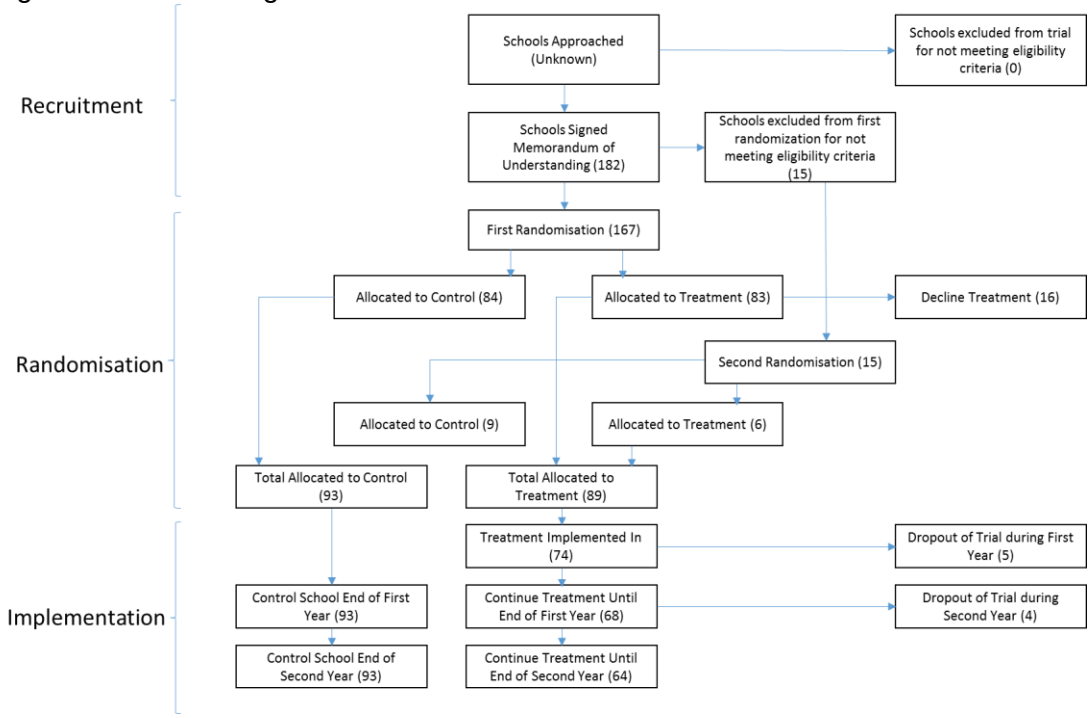


Follow-up

Figure 2 below shows a consort flow diagram detailing the randomisation and the follow up with schools. Note that it is impossible for treatment or control schools to drop out of the sample completely as we are still able to collect their test score data on the basis of the

Memorandum of Understanding. Some treated schools declined the program, or stopped implementing it during the two years of the trial. These schools will be considered as treated for the intention to treat effects.

Figure 2 Consort Diagram



Outcome measures

Primary outcome

The main outcome measure of interest is pupil academic outcomes as measured by Key Stage 2 (KS2) tests. Specifically our primary outcomes are KS2 combined average test scores across maths and reading for each cohort. These test scores will be obtained via the National Pupil Database (NPD). This database includes information on KS2 test results (for reading and Maths) as well as Key Stage 1 (KS1) test results for basic reading, writing and maths. We would primarily focus on KS2 test scores as the measure of pupil achievement for all pupils. Due to the change in KS2 assessment the KS2 composite English score will not be used as low correlations between reading and writing have been found in 2016.

Regarding the precise measurement, all KS2 scores will be percentalised at the national subject-cohort level to achieve comparability across subjects and years/tests. This will ensure that there is comparability in the distribution of test scores across schools and years. This is important given the changes that occurred to KS2 assessment during this period.

Secondary outcomes

Our pre-specified secondary outcomes are

- KS2 maths test scores
- KS2 reading test scores
- Teacher assessed KS2 Science level
- KS2 SPAG test scores

- KS2 writing test scores (SPAG). Note that from 2016 on, writing test are no longer externally assessed. Therefore, we will include writing test scores only for the years prior to 2016.

Analysis

Our preference is to conduct the analysis at the individual level. However, this might not be possible due to changes in the data availability and confidentiality policies by the Department for Education that occurred since the start of this trial. If the latter was the case, the analysis will be conducted at the school-cohort level. For the purpose of this document, the working assumption is that we get permission to link the collected UPNs to the PMRs of the NPD.

The analysis will follow EEF guidelines[2]. Our analysis plan in the protocol is different from the current guidance therefore both sets of estimates will be produced. All econometric analysis will be conducted with the Stata software package.

Primary Analysis

In accordance with the guidance the analysis of primary and secondary outcomes measure(s) will be undertaken on intention to treat basis meaning that all those allocated to treatment and control in the randomisation are included, even if they drop out of the treatment. This means that for all analyses the maximum 182 schools will be used. It should be noted that the schools that were left unmatched will not add to the estimation of the effect size as their average test scores will be absorbed by their pair fixed effect, which in this case will only represent this one school.

In a randomised design the effect size on the primary outcome will be calculated using OLS at the pupil level for increased power and to reduce bias, with clustering at the school level and reporting robust standard errors. Although a direct comparison of the means should be sufficient for determining the effect size, controlling for prior attainment (KS1) will increase the precision and reduce any potential biases. As explained in the guidance, adding covariates reduces the total variances to be explained. For comparability, additional covariates will not be included, with the exception of the pair-fixed effects. The estimated model will then include treatment status, KS1 attainment, and pairing as fixed effects using robust standard error. This will be estimated separately for Cohort 1 and Cohort 2.

When determining the effect size we will use the total variance, rather than the residual variance from the clustered model. Variations in a post-test outcome are due to different sources, which must be fully accounted for in a statistical model. For cluster randomised trials, the total variability can be decomposed into random variation between pupils (σ_i) and heterogeneity between schools (σ_s). Effect sizes for cluster randomised trial with equal cluster size and using total variance will be calculated as:

$$Effect\ Size = \frac{(\overline{Y_T} - \overline{Y_C})}{\sqrt{{\sigma_i}^2 + {\sigma_s}^2}} = \frac{\beta_{Treat}}{\sqrt{{\sigma_i}^2 + {\sigma_s}^2}}$$

We will calculate the effect of the program on the first cohort (C1) and the second cohort (C2) for the average KS attainment. See Table 2 below. An equivalent table for the secondary outcomes will also be produced.

[2] https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_for_EEF_evaluations.pdf

Table 2: Primary analysis

Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Cohort 1							
KS2 Average							
Cohort 2							
KS2 Average							

Protocol analysis

Given the nature of RCT will can use a basic differences method to estimate the impact of the programme. This will be clustered at the school level. Using Key Stage 2 scores means that we can use administrative data for both the treatment and control schools in the study which has many advantages. Firstly we have access to a long time-series of pupil attainment results in each school, which we will be able to compare any changes against. This means additionally we will also use a differences-in-differences approach to evaluate the impact of the programme on Key Stage 2 results, comparing changes over time in the results of treatment schools with those of control schools. This is dependent on the structure of Key Stage 2 assessment not changing significantly during the analysis period differentially between treatment and control schools. Any changes that impacts on both treatment and control schools would be absorbed in the year fixed effects. This will further increase the precision of the estimates. Secondly, as the compulsory test scores are collected centrally we do not need to be concerned with attrition due to testing in treatment or controls schools. This will ensure internal validity of the results.

We will estimate the impact of the programme on pupil outcomes using a difference-in-differences approach using the following model.

$$Y_{ist} = \alpha + \beta(Treat_s * Cohort_t) + \lambda X_i + \theta_t + \mu_s + \upsilon_{ist} \qquad (1)$$

The dependent variable Y will be the pupils' i KS2 test score from school s in year t. β represents the effect of the programme on pupils, $\mathrm{Treat_i}$ and $Cohort_i$ are indicator variables which will equal 1 for treated schools and students who took their KS2 examinations in the academic year corresponding to the appropriate cohort (C1=2014/15, C2=2015/16), and 0 in all other circumstances. We will also include a vector pupil characteristics X_{is} to take into account of their effects on test scores which will improve the efficiency of the estimations. The student characteristics will be; KS1 test scores in the relevant subject, gender, and FSM status when the student took KS1. Finally we include a set of school (μ_s) and year effects (θ_t) to control for any unobserved differences between schools, or across years, this will further improve the efficiency increasingly the likelihood of estimating significant results. We will cluster the standard errors at the unit of treatment which is the school (school cluster).

Note that this model will be run separately for each cohort. To avoid the potential issue of students in the pre-treatment years obtaining treatment e.g. the school provides the

treatment to year six pupils, the difference in difference analysis will include all years from 2008/9 up to the start of the trial (2012/13) as control years and the treatment year only. This means for C1 we will omit 2013/14 from the analysis and we will omit 2013/14 and 2014/15 for the analysis of C2. Joint effects of the average impact of the treatment across both cohorts will also be estimated by specifying a treatment dummy over two years. This will have more power, due to the increased number of observations, but will be more difficult to interpret as will represent the average impact on a cohort treated for one year, and another cohort for two years.

A key concern is that of non-compliance of treatment and control schools (never-takers and always-takers). Equation 1 represents the ideal situation with complete compliance and would provide an unbiased estimate of the Average Treatment Effect (ATE). In reality some schools that were asked to be in the treatment may not accept and therefore we will have to replace the treatment variable with one for intention to treat $Intend_i$. Estimating this equation will give as an unbiased Intention To Treat (ITT) effect. We define compliance of treated schools if we received confirmation from the school at the end of each academic year that they participated, and we received a UPN list of the students taught by a Lesson Study teacher (detailed below).

However policy makers are typically interested in the ATE, that is to say how big would the effect be on an average school, or student? To address this issue we will use an Instrumental Variable approach. This uses the original assignment lists to predict whether the school (student) will be treated. For this $Treat$ will represent if the school actually went through with the programme and $Intend_i$ is an indicator variable for if a school was randomly assigned to be treated. $Intend_i$ will be a strong predictor of whether or not the school was actually treated but as it is randomly assigned we know that it is independent of pupil outcomes and therefore if used as an instrument will account for bias due to non-compliance.

$$Y_{ist} = \alpha + \beta(\overline{Treat}_s * Cohort_t) + \lambda X_i + \theta_t + \mu_s + \varepsilon_{ist} \qquad (2)$$

$$\overline{Treat}_{st} = \alpha + \beta_0(Intend_s * Cohort_t) + \lambda X_{is} + \theta_t + \mu_s + \vartheta_{ist} \qquad (3)$$

We will also produce an ATE using variation at the student level. Note that although an entire school is allocated to treatment or control, not all students within a school year will necessarily be treated as some teachers in a cohort may not have been involved in the program. Therefore we have obtained the UPN of each student taught by a Lesson Study teacher, and so we know that have been exposed to the programme. For this student-level analysis $Treat$ will represent if the pupil actually went through with the programme and $Intend_i$ is an indicator variable for if a school was assigned to be treated. Again $Intend_i$ will be a strong predictor of whether or not the pupil was actually treated but as it is randomly assigned we know that it is independent of pupil outcomes.

$$Y_{ist} = \alpha + \beta(Treat_i * Cohort_t) + \lambda X_i + \theta_t + \mu_s + \varepsilon_{ist} \qquad (4)$$

$$Treat_i = \alpha + \beta(Intend_s * Cohort_t) + \lambda X_{is} + \theta_t + \mu_s + \vartheta_{ist} \qquad (5)$$

The results for these analysis will presented in Table 3 shown below. Again, an additional table for the secondary outcomes will be produced.

A note on the treatment dosage for cohort 2: Here, students should have been treated over two years. This could be important for the interpretation of the results and in particular for the comparison of effect sizes across cohorts. We expect that most children of cohort 2 have indeed been treated twice, which we will document once the data linkage has occurred.

Table 3: Protocol analysis

	Effect size								
	Intention to Treat (ITT)			School ATE			Student ATE		
Outcome	Difference in Difference	Hedges g (95% CI)	p-value	Difference in Difference	Hedges g (95% CI)	p-value	Difference in Difference	Hedges g (95% CI)	p-value
Cohort 1									
KS2 Average									
Cohort 2									
KS2 Average									

Interim analyses

Not applicable

Imbalance at baseline

We will use the administrative information from the NPD to test for balance on the school and student level characteristics. The summary statistics of the school will be from a baseline year of 2010/11, allowing us to compare outcomes before the experiment had started. We will present these summary statistics of the school and students as in Table 4 below.

Table 4 Balance of Observables

Variable	Intervention group		Control group	
School-level (continuous)	**n (missing)**	**Mean**	**n (missing)**	**Mean**
School Size				
% FSM				
Average KS1 Maths (2011)				
Average KS1 English (2011)	…	…	…	…
Average KS2 Maths (2011)				
Average KS2 English (2011)				
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for FSM				
Male				
Pupil-level (continuous)	**n (missing)**	**Mean**	**n (missing)**	**Mean**
KS1 Maths				
KS1 English				

Missing data

As we rely on NPD tests there should not be any missing data. If some individuals for whatever reason cannot be linked by the DfE we would refrain from imputations and drop these individuals from the analysis.

However, there are some schools in the analysis that did not have a year six and so had not tested students, or did not exist in 2010/11 because the new school is a result of an amalgamation. For these schools no baseline school level information is available and therefore will not be included in the balance tables, and will be counted as missing. No imputation planned regarding missing data.

On-treatment analysis

Assigned treatment dosage as in specification 1 above. When the analysis is carried out that the school level (main analysis, first half of protocol analysis) those in cohort 1 will have one year of dosage, those in cohort will have two years. The analysis at the individual level will count dosage depending on number of years taught by a lesson study teacher.

Secondary outcome analyses

The secondary outcomes will be assessed using the same methodology and approach as the primary outcomes.

Subgroup analyses

The NPD would also allow us the ability to measure outcomes for different types of pupil, and hence to detect heterogenous impacts. For example, the programme might improve teacher effectiveness at addressing the needs of FSM pupils, or previously low achieving students or those from poor backgrounds. In the protocol we specified five categorisations for subgroup analysis; 1) Free School Meal Eligibility; 2) English as an Additional Language; 3) Low Achievers (as defined by KS1); 4) Ethnicity; 5) Lesson Study specialisation of the school (Numeracy/Literacy). The majority of the schools did not record their specialisation of the Lesson Study programme and therefore we are unable to perform this subgroup analysis. Therefore this list has been reduced to the remaining four categories. Free school meal eligibility, ethnicity, and previously low achieving KS1 students, defined as those that were assessed at Level 1 or below.

The primary subgroup analysis will be producing intention to treat (ITT) estimates, using the initial allocation to treatment and control groups. The size of the differences of the effects between the groups will be determined by an interaction term of assignment to treatment and the characteristics of interest. The secondary subgroup analysis will be estimating the ATE at the student level (specification 3) because the subgroups are at the individual level. Again there will be two endogenous parameters (treatment and treatment*characteristic) and two instruments (intend to treat, intend to treat*characteristics), and the difference between the groups will be determined by the interaction term.

References

Raudenbush, S. W., et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]. Available from www.wtgrantfoundation.org.