Lesson Study
LSE
Richard Murphy

Evaluation Summary	
Age range	Primary (Year 4 and Year 5)
Number of pupils	c. 9,600
Number of schools	160 schools
Design	Randomised controlled trial with randomisation at the school level
Primary Outcome	Impact on KS2 scores

Lesson Study Protocol

Impact Evaluators: Richard Murphy, Amy Challen, Felix Weinhardt, and Gill Wyness.
Process Evaluators: Heather Rolfe, Anitha George

Abstract

Background: Teachers are now well recognised as the most important factor within schools for improving pupil achievement. However, an area which has hitherto received little research attention concerns the methods for improving the effectiveness of teachers. *Lesson Study* provides an excellent opportunity to examine the benefits of developing teachers through peer observations and structured feedback.

Study Design: A stratified randomised control trial evaluating the effectiveness of Lesson Study through the attainment of pupils in Key Stage 2 examinations. The unit of allocation is the school. These are recruited to the trial through 10 Local Authorities. The programme will be focused on teachers who teach year 4 and 5 pupils and will last for two years.

Discussion: As of May 2013, the first phase of recruitment of schools had begun. The randomisation will take place on the May 8th.

Background

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 the quality research that has been conducted on teacher development, knowledge based training has typically been found to be ineffective, whereas programmes that involve teacher observations that are embedded into the school alongside effective feedback have been found to have effects (Taylor and Taylor, 2013).

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.

Lesson Study is will be one of the largest randomised trial to date focusing on the effectiveness of this form of teacher development.

Our evaluation strategy involves randomly assigning the Lesson Study programme to 80 schools across ten local authorities (Cambridgeshire, Central Bedfordshire, Devon, Lancashire, Liverpool, Manchester, Peterborough, Plymouth, Stockport and Torbay). The programme will be focused on teachers who teach year 4 and 5 pupils and will last for two years. We will use literacy and numeracy Key Stage 2 test scores for the main outcome measures.

Methods and design
Research Questions
There are three main aims to this evaluation:
1. To assess whether the attainment (standardised test scores and Key Stage 2 results) of pupils in the intervention schools improves relative to that of similar pupils in the control schools.
2. To assess the impact on pupils by subgroup relative to that of similar pupils in the control schools.
3. To assess whether the impact of the programme increased over a two year period.

Design & Participants
The evaluation will take the form of a stratified randomised control trial, to estimate the effectiveness of Lesson Study through the attainment of pupils in Key Stage 2 examinations. The unit of allocation is the school and 10 local authorities will be involved (Cambridgeshire, Central Bedfordshire, Devon, Lancashire, Liverpool, Manchester, Peterborough, Plymouth, Stockport and Torbay). However, the programme is capacity constrained by the volume of trainers that Lesson Study can provide in each local authority. Therefore, a provisional quota of schools that can be treated is in place in each Local Authority as follows:

Cambridgeshire 10
Central Bedfordshire 10
Devon 10
Lancashire 10
Liverpool 10
Manchester 5
Peterborough 10
Plymouth/Torbay 10*
Stockport 5
   *Plymouth/Torbay will share 10 schools

Schools will be recruited by Local Authorities using the following method:
Local Authorities will contact 4 times the quota of schools in each LA and ask for expressions of interest in the programme Preference will be given to schools with high proportions of Free School Meal Eligible (FSME) children and those with a two form entry or less. The Local Authorities will not approach all applicable schools at this stage.

Among those who express an interest, 80 will be randomly selected into the treatment, stratified by Local Authority. We will be using pairwise to ensure similarity of characteristics between treatment

and control. These will be the proportion of FSME pupils and average Key Stage 1/2 attainment. To ensure only 80 schools receive the treatment the randomisation will work in the following way: if 160 or fewer schools express an interest similar pairs will be formed of which one will be selected into the treatment, as per standard pairwise randomisation. If there is in excess of 160 expressions of interest, the appropriate pairs will be formed and then 80 of the pairs will be randomly selected (stratified by Local Authority) to take part in the pairwise randomisation. Those selected as treatment schools will then be approached to confirm that they will take up the programme. Schools that drop out at this stage will remain in the treatment group for the purposes of calculating intention to treat effect. If there are now fewer than 80 schools who are going to receive the treatment we will randomly select from the remaining pairs who haven't yet had pairwise randomisation. For example if there are 200 expressions of interest, we will form 100 pairs of schools. We will then randomly select 80 pairs, stratified by local authority, to go through the pair wise randomisation. Once selected if any of the designated treatment schools then immediately decides to drop out the short fall will be made up from the 20 remaining pairs.

A second phase of recruitment will be implemented if there are less than 80 schools designated to receive the treatment. In this phase the Local Authorities invite previously un-approached schools. The randomisation into the treatment will then follow the same method as above. This will ensure that 80 schools participating in the programme. The schools that expressed an interest but who were not selected will act as control schools.

Analysis plan

Given the nature of the 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 in the in betweening period.

The other advantage of this approach is that it is 'light touch' – i.e. we do not need to directly involve control schools in any element of the study allowing us more treatment schools, relative to a rolling treatment trial. Finally, as the test scores are collected centrally we do not need to be concerned that treatment or controls schools will attrit from the study. This will ensure internal validity of the results.

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

$$Y_{ist} = \alpha + \beta(Treat_i * 2013_t) + \delta Treat_i + \lambda X_{is} + \theta_t + \mu_s + \upsilon_{ist} \qquad (1)$$

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, $Treat_i$ and 2013_i are indicator variables which will equal 1 for treated schools and for academic years from 2013/14, and 0 in all other circumstances. δ accounts for average differences between the control and treatment schools over the entire period (including pre-treatment). If the randomisation has been well implemented we would expect . $\delta = 0$. We will

also include a vector of school and pupil characteristics X_{is} to take into account of their effects on test scores which will improve the efficiency of the estimations. 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 but common to all schools, 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. This will be estimated over the 2001-2016 period.

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 us an unbiased Intention To Treat Effect (ITT), which is likely to be the first model estimated unless all schools offered the scheme took it up.

However policy makers are typically interested in the ATE, that is to say how big would the effect be on an average school that took up the programme. To address this issue we will use an Instrumental Variable approach. This uses the original assignment lists to predict whether the school 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 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.

$$Y_{ist} = \alpha + \beta(\overline{Treat_i} * 2013_t) + \delta\overline{Treat_i} + \lambda X_{is} + \theta_t + \mu_s + \varepsilon_{ist} \qquad (2)$$

$$\overline{Treat}_{st} = \alpha + \beta(Intend_i * 2013) + \delta Indend_i + \lambda X_{is} + \theta_t + \mu_s + \vartheta_{ist} \qquad (3)$$

Outcome Measures
The outcome measures for the programme will be the Key Stage test scores of year 6 pupils. Specifically we will analysis the effect on:
- Average KS2 test scores
- Maths KS2 test scores.
- English KS2 test scores

Furthermore there are three secondary outcomes
- Reading KS2 test scores
- Writing KS2 test scores
- Science KS2 test scores

These will be obtained from the National Pupil Database, which centrally records achievement of all pupils.

Subgroup analysis
There will be additional subgroup analysis by the following groups of pupils:
- Free School Meal Eligible
- English as an Additional Language

- Low Achievers (as defined by KS1)
- Ethnicity
- Lesson Study specialisation of the school (Numeracy/Literacy)

Sample size calculations

An important concept when designing an evaluation is statistical power. Increasing the number of treatment and control schools increases the statistical power which is the likelihood of determining a significant result if it exists. Using a conservative assumption of a cohort size of 30 pupils in each school cohort and using the stated impact of 0.2 s.d our method of would require 60 schools (30 treatment, 30 control) to be confident of finding an effect. However, using an even more conservative impact of 0.1s.d., which is still considerable in the education literature, we would require closer to 160 schools to be confident in finding an effect (See Figure 1). Given our recruitment and selection strategy we will have 80 treatment schools and at least 80 control schools meaning that the evaluation will have a sufficient power that we would be confident of finding an impact if there is one.

Figure 1: Power graphs, with 30 pupils per school and 5% significance
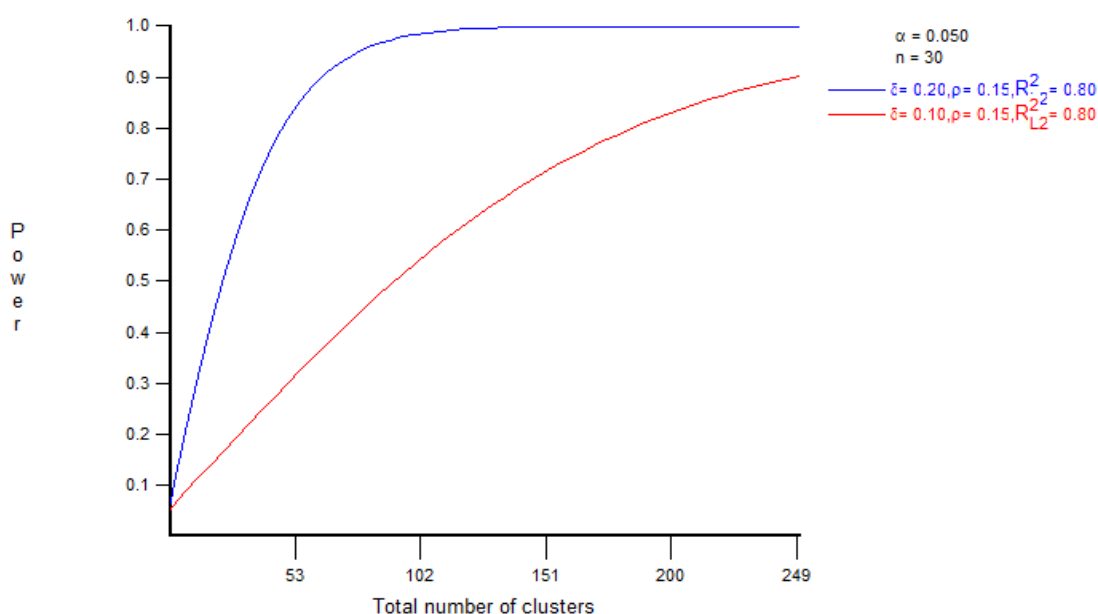
Process evaluation methods

Aim of the process evaluation

The aim of the process evaluation is to understand teachers' and schools' perceptions of the programme and identify factors affecting its successful implementation, with a view to wider roll-out. The evaluation will focus on processes surrounding its introduction and embedding, but will also identify issues relevant to its impact. We will aim to identify features which contribute to successful implementation, including practicalities and of factors such as staff confidence in the approach and senior management commitment. Qualitative research will also be aimed at bringing greater clarity to the quantitative research findings and understanding the reasons for and mechanisms behind the estimated impacts.

Design of the process evaluation

To ensure efficiency and prevent duplication and excessive demands on participating schools, the process evaluators will work closely with the other project partners: Lesson Plan; the project manager at Edge Hill University (to be appointed) and the quantitative impact team. Edge Hill will be conducting an internal evaluation, and we intend to exchange research materials within the team. Fieldwork visits will be coordinated by Edge Hill to avoid multiple contact with schools resulting from internal and external evaluation and project visits. Regular updates will be arranged between the all parties to identify emerging issues of relevance to the development and implementation of the programme and to ensure joined-up working.

To distribute the demands which the project and evaluation make on schools, we will conduct the process evaluation in the South West and East of England, so that Edge Hill University can focus its internal evaluation on North West schools (to be confirmed). Edge Hill University will advise NIESR on sampling, to obtain a range based on their team's knowledge of the school characteristics.

The process evaluation will capture participants' experiences of the pilot programme at 2 main stages: during training and during implementation and application of Lesson Study at classroom level. In addition to understanding teachers' experiences of the programme, it will be important to obtain senior leaders' perspectives and experiences. This is because issues such as commitment to the programme, through belief that it will improve the quality of teaching, are likely to influence the effectiveness of the programme, and have implications for wider roll out. The research will consist of the following components:

- Attendance at the introductory conference in September to evaluate training
- Analysis of training and lesson delivery materials (produced by Lesson Study)
- Attendance at all stages of the training programme in one region
- Visits to 10 schools (5 schools in 2 regions) to interview staff and senior managers
- Face to face and telephone meetings with Lesson Study and Edge Hill University
- Analysis and report writing

MANAGEMENT AND ORGANISATION

Personnel

Impact Evaluators: Richard Murphy, Amy Challen, Felix Weinhardt, and Gill Wyness.

The Impact evaluation team will based at the LSE and will be led by principal researcher Richard Murphy. The team consists of Ms Amy Challen, Dr Felix Weinhardt, and Dr Gill Wyness. All the evaluators have expertise in the education literature. Richard Murphy has written numerous papers on teacher effectiveness. Amy Challen and Gill Wyness has worked on other field RCTs.

Process Evaluators: Heather Rolfe, Anitha George

The Process evaluation team will be based at NIESR and led by Dr Heather Rolfe (Principal Research Fellow) who will work with Anitha George (Research Fellow) plus some additional input from a Research Officer.

Timeline

April 2013

CEP- Lesson Study: Agreement on methodology between evaluators and the Lesson Study partners; e.g. list of potential schools that could be approached

April 2013

Lesson Study: Liaise with local authorities and schools to recruit schools into the programme

May -2013

CEP- Edge Hill: Get feedback from schools that are interested in the programme

May-2013

CEP: Randomly assign schools to treatment from the interested eligible group.

CEP-Lesson Study: If target number of treated schools is not met imitate second phase of recruitment.

CEP: Randomly assign schools to treatment from the interested eligible group.

September 2013-June 2014

Edge Hill: Implement Lesson Study in selected schools.

NIESR: Evaluate training programme

NIESR: Visit sample of ten schools to interview staff and senior managers

NIESR: Discussions with Lesson Study and Edge Hill university

NIESR: process evaluation report

December 2015

CEP- First release of KS2 results in NPD by DfE for first cohort of pupils

February 2016

CEP: Write interim report for first cohort of pupils

December 2016

CEP- Frist release of KS2 results in NPD by DfE for second cohort of pupils

February 2017

CEP: Write final impact.

Risks

Non-compliance: As is common in Randomized control trials, we expect there to be some non-compliance. This could take two forms – i) schools assigned to the treatment group do not take up the programme (never-takers) and ii) schools assigned to the control group take up the programme (always-takers). We have proposed a method to deal with this non-compliance, using an IV strategy. However for this to be effective, we will need to communicate closely with Lesson Study to receive on going information on the compliance or otherwise of schools in both the treatment (did they take it up?) and control groups (are other schools taking up the programme?).

The risk of non-compliance also applies to the process evaluation. The 'light touch' design should help to minimise this risk, as will flexibility in the timing of visits and interviews on the part of the research team. One consequence of high levels of non-compliance is that the sample of schools visited is unusual in some way, for example they are more committed to the programme. The process evaluation team will work closely with Edge Hill University to ensure that such bias does not occur.

External validity: We hope that the results of the evaluation would be generalizable to the schools population with similar characteristics to those schools selected for the experiment . The current selection criteria mean that the majority of schools will have high proportions of FSME pupils, which means that we would only be able to make strong statements about other similar schools. However, as the Local Authority has some discretion in which schools are approached it is hoped that there will be a broad range of schools included in the programme.