Target oriented lesson study (TOLS) Combining Lesson study with an integrated impact evaluation model

Abstract

This paper presents the learning gained from two projects in London from 2013 to 2015, both of which primarily involved improving teacher pedagogy and pupil learning outcomes through the promotion of lesson study. Students from under-achieving or vulnerable groups were targeted as the source of the interventions. One project was aimed at mathematics teaching in primary schools, while the other involved both primary and secondary school teachers and were targeted at improving outcomes for students in either mathematics or literacy. An impact evaluation approach was adopted that had been previously used at the authors' university department to evaluate other professional development programmes and this was adapted specifically to use in lesson study cycles. The approach combined the ideas of i) 'starting with the end in mind', i.e. targeting the impact on pupil learning and changes to practice, ii) using Guskey's multi-level logical chain of impact to guide the use of data collection tools and iii) an 'improving rather than proving' approach to evaluation that sought to maximise the impact of lesson study on practice. These programmes showed that the lesson study cycles had significant impact on teacher learning, and where organisational cultures and structures were attended to teachers were successful in developing new knowledge and skills that led to explicit gains in pupil outcomes after one year of the project. The methodological approach to impact evaluation shows promise in that it enabled teachers to focus their minds on what they were trying to achieve with pupils; helped to maximise the effect of lesson study cycles on their pupils' learning and also showed the key factors that led to the success of lesson study as an approach to professional development in schools. A conceptual model of 'target oriented lesson study (TOLS) is suggested and areas for future development and research are identified. One particular challenge is to link changes in teaching strategies to specific qualitative pupil learning gains and we also discuss the 'thorny' issue of quantifying changes to learner outcomes.

Introduction

Aims:

This article sets out the learning from two major projects in London in which lesson study was used as the primary vehicle to increase teacher knowledge and skills in order to improve pupil outcomes in numeracy and literacy. The aim of this paper is to describe a model of evaluation that shows promise in relation to the use of lesson study cycles by teachers in both primary and secondary school settings.

We set out to show a methodological impact evaluation model that can a) show various levels of impact based on Guskey's model of evaluation of professional development programmes; b) that can help focus teachers' minds on the impact they are intending to have on their pupils, and by corollary on the practices that lead to these changes and c) combining a range of measures, both quantitative and qualitative, that show the impact of lesson study on teacher and pupil learning. Finally, we hope to show how evaluation, when 'built-in' from the start, has the potential to build focus, coherence and thus maximise professional learning and knowledge development in ways that help maximise outcomes for pupils.

We outline previous research, showing how lesson study has been evaluated and identify gaps in their methodological approaches. We present our own conceptual model and revise this in the conclusion in light of the learning gained from these two projects. Throughout, we try to show how there needs to be a concern for a parsimonious and practical approach to evaluation that schools can use, while being sufficiently rigorous to be able to identity the features of the programme that led to its successes and failings. The paper also invites further debate and research on ways in which lesson study can be evaluated more meaningfully, without disrupting the principles of the approach as a form of inquiry and professional pedagogy.

Context

This article summarises the learning from two London Schools Excellence Fund (LSEF) funded projects that ran from 2013 to 2015. LSEF funds came from the London Mayor's Education Programme and had a budget of over £20 million, most of which originated from the Department for Education. The fund itself was established to improve the quality of teaching in the capital, and over 100 projects have been awarded funding, covering 18,000 teachers across the Greater London metropolitan area[1]. The two projects from which the empirical data have been taken were called "Ascend" and "Lambeth Connecting Knowledge (LCK)"; each were concerned with improving teaching practice, and thus outcomes for learners, primarily via engaging teachers in Lesson Study. Students from under-achieving or vulnerable groups were targeted as the source of the interventions. LCK[2] was aimed at mathematics teaching in primary schools, while the Ascend project[3] involved both primary and secondary school teachers and were targeted at improving outcomes for students in either mathematics or literacy.

Both projects were supported by staff working from the London Centre for Leadership in Learning at UCL Institute of Education, primarily by two of the authors of this paper. This involvement included co-planning of the project logistics, facilitating lesson study training and conducting process and outcome evaluation. In both cases, the projects sought to spread the use of lesson study across schools in the locality. These were either within a Teaching Schools alliance (akin to Professional Development School networks in the USA) or in other forms of local collaboration.

[1] https://www.london.gov.uk/what-we-do/education-and-youth/improving-standards-schools-and-teaching/london-schools-excellence?source=vanityurl

[2] http://www.rosendale.cc/learning/school-curriculum/

[3] http://www.ascendlearning.org.uk/home

Introduction

Lesson study in the English context

Lesson study has become increasingly popular in the last decade or so as a form of professional development in English schools (EBT article ref?, Lewis, 2006, Barber, 2007). The recent national focus on the need for more effective teacher professional development models that draw on research into effective learning (Hallgarten 2014, Sutton Trust 2015) have also led to a surge of recent interest.

As such, it was not surprising that a number of the new 'mayor's fund' (LSEF) projects involved proposals by school leaders to raise standards in this way. The global race to increase pupil attainment and to tackle issues of inequality in outcomes has led to an increase in what Steiner-Khamsi (year) describes as 'travelling reforms'; reforms borrowed from other nations that are perceived to be using models of excellent practice. The Far East, and Japan, Singapore and China in particular have seen several key aspects of their educational systems lauded internationally (chapter from Moore).

One of the dangers of borrowing ideas from elsewhere is in the 'dilution' of those aspects of practices that make them particularly successful in originating countries. On the other hand, there is something to be said for adapting approaches to best fit the national and local contexts. Thus, acting as HE facilitators, programme co-designers and evaluators, one of our concerns was to ensure fidelity to the key features of lesson study that would need to be incorporated into these programmes, while allowing for flexibility and fitness for purpose**.** While not the place in this article to elaborate all the distinctive features of 'proper' lesson study (see Takahashi and McDougal, 2015 for a more detailed analysis), we summarise the essential features here as involving:

- Identification of a research focus through an analysis of 'data' (in its most inclusive definition);
- collaborative lesson planning by a group of teachers;
- a research lesson, in which one member of the planning group teaches the lesson and all teachers in the group observe the lesson;
- a post-lesson discussion to draw out learning in relation to the research lesson.

(e.g. Stigler and Hiebert, 2009)

Other authors have stressed the importance of:

1. '*kyozai kenkyu', i.e.* the initial exploration of what is known about the research theme through a close examination of relevant material: curricular material, manipulatives, relevant research material (Murata (Hart, 2011) and Fujii (2013).
2. Ownership of the lesson; the lesson plan is seen as a group product, with shared ownership across the group of teachers (Stigler and Hiebert, 1999).
3. The role of the *kochi* (expert other) (e.g. Fernandez and Yoshida, 2004) in pulling together ideas from other lesson study groups and building on learning.
4. Revising and re-teaching the lesson. Some see this as an important part of lesson study (e.g. Stigler and Hiebert 1999, Lewis 2009) while others feel that this betrays the principle that each lesson is designed with one particular learning context in mind (e.g. Fujii, 2013)

Our lesson study cycles varied slightly between the two projects, but according to the above features, we used an approach faithful to all the above 'essentials'. The one caveat being that the

research focus was sometimes generated entirely by the teacher while in most others, the theme was an agreed one and then specifics as they applied to each focus teacher's classes were identified. Regarding points 1-4: in LCK, there was an element of the project that involved separate professional development around specific pedagogy of areas of mathematics (sarah – can we be more specific?). However, for Ascend, the extent to which an exploration of the evidence in relation to the theme occurred is somewhat unknown, but this was not stressed in our training or that of one of the school leaders who was instrumental in explaining lesson study to the schools. For 2, shared ownership of lesson planning was the intention of the lesson study 'trios' but in some instances the 'focus teacher' (the one taking the class) felt more that they took the lead and 'owned' the lesson than in others. This was influenced by the extent to which the focus teacher was identified (by the project and colleagues) as a 'lead' teacher and also the perceived expertise of the others in the group. In some cases, the level of trust in newly formed trios made it difficult for focus teachers to take on board new ideas for teaching their pupils, in particular in lesson one of the cycle. In most cases this changed and by lesson two, focus teachers were able to share the planning more openly with their colleagues. Ascend did not use a 'Kochi' role, rather there were guidelines on leading and facilitating post-lesson discussions that were rotated between participants as they each carried out a cycle as focus teacher. However, for LCK, an experienced external consultant (in mathematics and lesson study) was used for one cohort of teachers in each year; this kochi role then, to an extent, became the first cohort's duty in year 2 as they led a lesson study trio in a neighbouring school. Finally, the model of teach and re-teach was used in both projects. This was partly due to the fact that by having a second lesson in a cycle with the same teacher, we could track the impact of this process (usually over 2 or so weeks) on this teacher's learning and changes to their pupils. This was a subsequent (i.e. not the same) lesson, taught to the same pupils. This gave a sense of 'micro-impact' and some detailed tracking of pupil outcomes that were coded on a survey. This is summarised in the table below:

Lesson study element	**Ascend**	**LCK**
Identification of a research focus	✓	✓
collaborative lesson planning	✓	✓
research lesson	✓	✓
post-lesson discussion	✓	✓
kyozai kenkyu	✗	partial
Ownership of the lesson	✓	✓
Use of a kochi	✗	✓ (for new cohorts of future lesson study leads)
Revising and re-teaching the lesson	✓	✓

Measuring the impact of lesson study as professional development

Our approach to measuring the impact of lesson study was informed by previous practices and literature emanating from the London Centre for Leadership in Learning at UCL Institute of Education. The impact evaluation approach has been articulated formally in an article by two colleagues in relation to professional development in education (Earley and Porritt, 2014). Below we briefly re-state some of the central arguments and principles expressed in this article as well as a wider review of the literature. We then apply this model to lesson study in particular.

The first principle is that evaluating professional development (PD) is an important part of a cycle of work in schools in which a 'logical chain' (Ofsted 2006) is established between the PD work and its eventual effect on student outcomes. As such,

"the principles and approaches of impact evaluation are better understood and established from the outset of a development activity, rather than as an after-thought or an accountability measure"

(Earley and Porritt, 2014 p.113)

Previous reviews of the literature had noted that there was insufficient work done on evaluation of PD in schools and that the link to pupil outcomes was frequently missing or in-precisely articulated (Pedder et al. 2010, p.18). It is thus important,

"that the anticipated learning outcomes are explicit from the beginning" (Porritt, 2005).

In order to track impact towards the anticipated outcomes, the first step is to be very clear and explicit about the baseline measures in relation to these outcome targets in terms of both pupils and teachers' practices. By focusing more on the desired impact the evaluation process focuses much more on the difference the development activity has had, particularly on pupils, and less on the details of the activity itself. In other words it evaluates the impact on pupils rather than the impact of x activity. The effect of these principles is to avoid using school leaders, or teachers 'favourite' improvement processes and to find out how effective they are, but rather to focus on what can be done to improve pupils learning (however this is achieved).

Much of these principles have been expressed in an OECD working paper (Earl and Timperley, 2015) on evaluative thinking for successful educational innovations. The paper proposes a complex interplay between the innovation and evaluation in which, to ensure success, the latter is integral to the former:

"Evaluative thinking contributes to new learning by providing evidence to chronicle, map and monitor the progress, successes, failures and roadblocks in the innovation as it unfolds. It involves thinking about what evidence will be useful during the course of the innovation activities, establishing the range of objectives and targets that make sense to determine their progress, and building knowledge and developing practical uses for the new information, throughout the trajectory of the innovation" (p. 8).

In line with our thinking about evaluating the process of lesson study, we wanted to build evaluation into the cycle of professional learning in order to provide both the necessary focus on empirical evidence, feedback loops to maximise the effectiveness of the approach and a laser-like focus on how this was affecting desired outcomes for pupils.

Another important dimension to the work of the Centre has been to focus on the importance of leadership of PD. We know that by having school leaders closely involved in PD, including modelling

and participating in it and taking a lead in enabling the conditions for effective PD are in place, is arguably the most important feature of effective leaders (Robinson, year). Earley and Porritt (2014) have articulated the challenges for leadership in nine key features and three dimensions (see below). In terms of our lesson study projects, there were a number of key areas to address, such as familiarity with the process, separation of accountability and performance-management driven observations from lesson study observations and the creation of optimal conditions that allowed for teachers to work collaboratively. There was also strategic leadership taken in both projects that brought together neighbouring schools in lesson study groups and in sharing practices, these were particularly undertaken by a Deputy Headteacher in the Ascend project and by a key Headteacher in LCK.

(Earley and Porritt 2014, p.120)

To sum up the Centre's approach to evaluating PD:

1. Evaluation should be built-in at the start not bolted-on later
2. Anticipated, desired impact on pupil learning should be articulated from the outset
3. Baseline evidence needs to be clearly articulated and made explicit
4. Leadership of PD is crucial to its success
5. The logical chain needs to be explored in relation to PD (not missing out the steps that lead to changes to pupil outcomes)

In the LS projects that we were involved in we addressed the above issues by:

1) Being involved, as much as possible, as HE advisers in the co-creation of the programmes and in its modifications and developments for year two and beyond; 2 and 3) Giving clear guidance on how to write qualitative statements and to set quantitative measures of baseline and impact for teachers' own practice and, crucially, students' learning. This was achieved using an 'impact frame' sheet, which we return to later; 4) By involving school leaders in conferences and briefings, and by providing lesson study leaders with guidance to give to school Headteachers and 5) breaking down the impact evaluation into levels, based particularly on work by Guskey (2005).

To expand further on this fifth aspect, Guskey (2005) describes five different levels at which the effectiveness of professional development can be measured. Based on Lewis (2000) and Fernandez and Yoshida (2004) and using Guskey's categorisations, we might identify several anticipated outcomes for lesson study which are relevant to an English context.

1. *The organisation's professional development model*

The structure, time, resourcing of the school's professional learning programme would need flex in order to accommodate lesson study. Cultural attitudes towards professional learning would be likely to shift in relation to: the role of peer-to-peer learning, teacher ownership of learning, lesson observation as learning not performance. Schools might begin to access networks beyond the school, as per the Japanese model.

2. *Teachers' reactions*

Teachers' attitudes to and enjoyment of professional learning might improve.

3. *Teachers' professional learning*

There could be an impact on subject knowledge, pedagogical content knowledge and confidence

4. *Teacher use of new knowledge and skills*

Teachers' newly acquired confidence, subject knowledge and pedagogical content knowledge could lead to changes in practice.

5. *Pupil learning outcomes*

Changes in teachers' practice might lead to improved attitudes to learning and progress for pupils, in terms of evidence from written work and assessment data.

While some authors have suggested that the domains that lead to professional growth are inter-connected and non-linear, unlike Guskey's model (see Bubb, 2013, Clarke and Hollingsworth, 2002), we nevertheless found the five levels to be a useful way of expressing the various stages of the logical chain towards (hopefully) improved pupil outcomes. This helped us communicate these ideas to participating teachers and to frame the various evaluation and impact tools that we had designed. The levels that Guskey outlines, also provide an interesting way to frame a systematic review of the literature, to which we turn below:

Measuring the impact of lesson study, a systematic review of the literature

This literature review focuses on empirical studies that have looked at the impact of lesson study on practising teachers. 186 studies involving lesson study were located as having been published between 2005 and 2015. An additional 25 were located through checking other recent systematic reviews of lesson study literature (Saito 2012, Wong 2014 and Xu and Pedder 2014). Of 181 rejected studies, 29 were about pre-service teachers, rather than in-service teachers and another 4 were about the involvement of additional school staff in the lesson study process (e.g. teaching assistants, librarians); 37 studies were descriptions of the lesson study process, either in the form of analyses or practical handbooks for practitioners; 14 focused on the challenges of adapting lesson study to different international contexts (beyond Japan); 10 were critical essays, setting out the authors' opinion of the lesson study model. Of the remaining studies, 30 were not practising lesson study as defined in the above literature review for various reasons, including:

- There were no live observations of lessons (for example, several studies use video instead of live observation).
- The form of lesson study used was a hybrid of other professional development models (for example, Learning Study, Design Study or Action Education as identified by Xu and Pedder (2014).
- Lesson study was used alongside other professional development models in a blended programme of support (for example, coaching, team teaching), meaning the direct impact of the lesson study element of the programme was unclear.
- The lesson study process focused on the polishing of one 'perfect' lesson.
- Lessons were followed by 'feedback' as per a standard classroom observation.

18 had a very narrow focus on the development of a specific area of subject content or a specific teaching skill, often through the 'polishing' of one lesson or a sequence of lessons (which could also be seen as contrary to the definition of lesson study as seen above); 13 lacked a clear methodology or empirical data. Others focused on an in-depth analysis of one aspect of the process, for example, the role of the lesson study facilitator (Vaughn, S. 2012), an analysis of teacher talk in lesson study (Murata, 2012, Clevenger 2009), an analysis of teacher reflection (Dudley 2013, Suratno 2010, Warwick 2016).

This left 27 studies which described impact on one or more of the Guskey aspects identified above. Of these 27, studies varied in their fidelity to the Japanese model in terms of the role of the *kochi* with only 19 specifying some involvement of a *kochi* type role. 13 of these studies took place in the USA and 7 in the UK, with others taking place in Holland, Indonesia, Japan, Ontario, Philippines and Singapore. In terms of sample sizes, numbers of teachers ranged from 4 to 83 (with three studies unclear about the exact total number of teachers involved). Of the 27 studies, 15 involved less than 10 teachers and only 2 studies involved more than 50 teachers. The average number of teachers involved in the studies was 17. Numbers of schools involved ranged from 1 to 9 (with seven studies unclear about the exact total number of schools involved). Of the 27 studies, 18 involved 5 schools or less. 13 studies involved only 1 school. (summary table here Sarah?)

In summary then, the studies are largely small-scale case studies involving small numbers of teachers and schools.

All 27 studies measured and/or analysed changes caused by lesson study in relation to teachers' learning, including one or several of the following aspects: subject content knowledge, pedagogical content knowledge, teacher confidence. Of these, only 10 studies considered the impact the lesson study process had had on pupil learning (attitudes and/or progress) although the majority of these studies relied on teacher reports of improvements to learning, rather than standardised measurement tools such as questionnaires or tests. Similarly, only 18 studies considered whether teachers had changed their practice in response to their involvement in lesson study. The literature describes lesson study as a 'comprehensive system for teacher learning' (Lewis, what is), and yet only 12 of the studies looked at impact on teachers and pupils beyond the lesson study cycle itself. This means that 56% of studies focused on analysing and measuring improvements during the research lessons and there was no evidence of transfer of learning to teachers' general classroom practice nor of impact on pupil learning more broadly. 18 studies analysed/measured teachers' reactions to and enjoyment of the lesson study process. Only 5 studies considered the impact the lesson study process had had on the schools' professional development model.

Overall, none of the studies we reviewed were able to capture the entire 'logical' chain as set out against Guskey's five levels, applied to lesson study programmes. The sample characteristics in much of the literature is either insufficiently clear or uses very small samples of teachers, schools or is limited to specific aspects of lesson study. One current study that we are aware of, led by Edgehill University does intend to measure impact on pupil learning, as well as to 'understand teachers' and schools' perceptions of the programme and identify factors affecting its successful implementation' (Education Endowment Foundation, 2016). However, this has yet to report its findings.

Below we set out an approach that covered all aspects of this lesson study process, based on a large sample of teachers and schools. We outline our evaluation tools and approaches, suggesting a methodological model for impact evaluation in lesson study programmes and avenues for future research based on this model.

Methodology

Lesson study cycles

Both projects worked with similar lesson study cycles of enquiry. In these, one teacher would be the focus of a cycle and worked alongside colleagues (usually two others, and in some cases an external facilitator) to co-plan a lesson, then the focus teacher would take a class with their pupils and colleagues would observe the learning of case study pupils (between 3-6), followed by a de-briefing session. In the post lesson discussion, the lessons learned would be used to plan a second lesson and this would follow the same pattern.

Data collection

The data is taken from the second year of the intervention and combines the results from Ascend and LCK in the 2014-15 school year. In the first year (2013-14), evaluations through surveys, interviews and pupil data were conducted and these led to changes in the programme in year two. Evaluation tools were also refined and improved in order to meet the needs of the schools, teachers and the requirements of the funders, who required detailed written reports. The pupil cohorts are also separate from year one to year two of the projects, thus here we report only on the pupil outcomes from the 2014-15 cohorts.

Sample characteristics

Teachers, schools and pupils

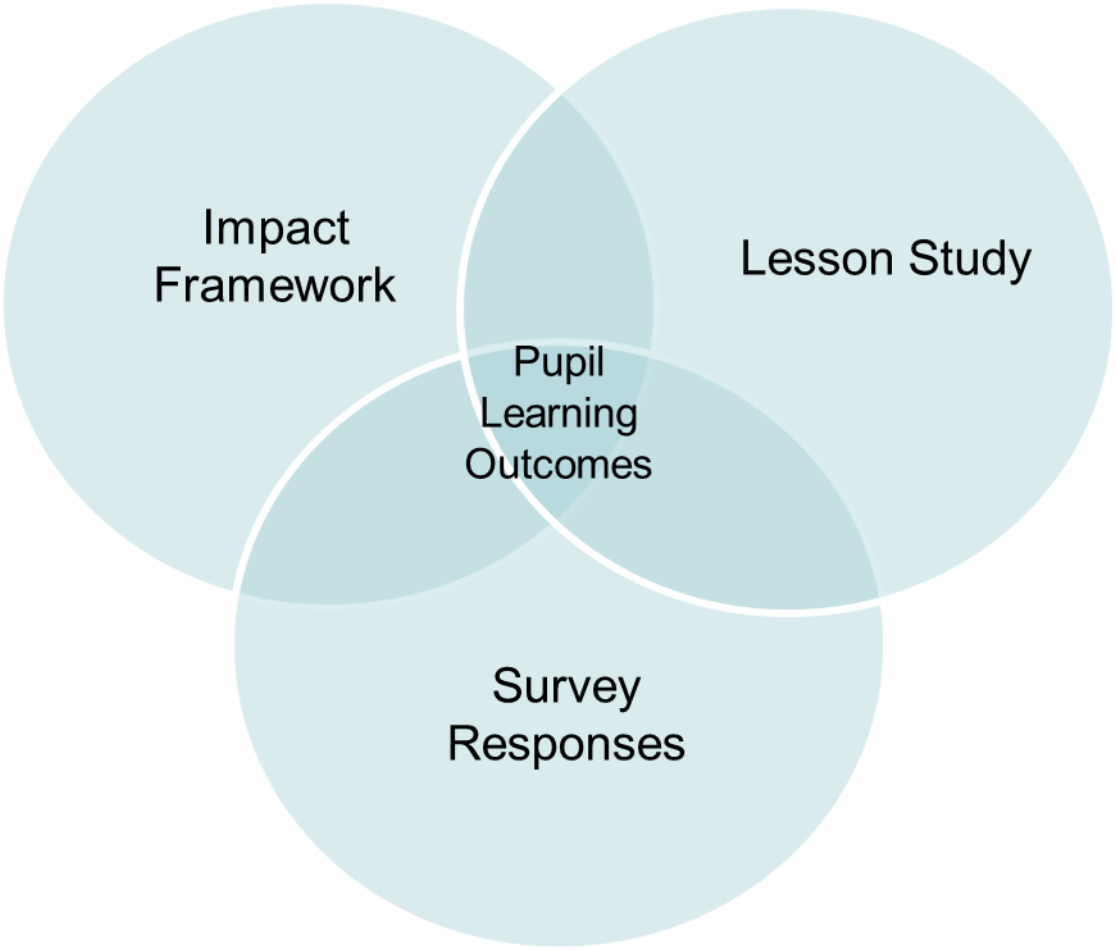
In year two of the projects there were 133 teachers (38 from LCK and 95 from Ascend) and 33 schools: 26 Primary and 7 Secondary. All 14 schools in LCK were Primary phase. Teachers were selected to take part in the projects by senior leaders at the school using a range of criteria and the project was aimed to help pupils who were disadvantaged and/or behind their peers in terms of progress. With most teachers selecting 3 target pupils and some up to 6, this makes a total of over 400 pupils. In practice, due to the challenge of ensuring completion of evaluations by teachers from a range of schools, in particular the Ascend project, we had some missing data. The summary table of data collection tools shows the varying response rates for each measure. In the Ascend project, 70% of target pupils were classed as 'White British' and the next largest groups were either Black African (7%) or Black Other (9%). LCK pupils were more diverse, as might be expected of their inner-city location, with Black Caribbean being the biggest single ethnic group at 22% and White British second at 17% and a wide range of other minority groups. A high proportion of pupils were eligible for free school meals (65% in Ascend and 44% LCK) and many were in receipt of support for special educational needs (27% Ascend and 12% LCK). Also, 25% of Ascend pupils were, or had previously been fostered or adopted (none for LCK).

Impact evaluation tools

As mentioned in the introduction, our evaluation methodology was influenced particularly by Thomas Guskey's approach to establishing the logical chain between teacher learning and learner outcomes. This meant that our evaluation tools needed to capture: initial participant (teacher) reactions; professional learning; organisational support and change; use of new knowledge and skills and finally, pupil learning outcomes.

From the perspective of the teachers involved in lesson study trios, there were three elements to our approach. There were: i) setting specific final impact goals for case pupils and teaching practice, ii) the planning and conduct of lesson study cycles and iii) evaluating LS cycles as the focus teacher and also overall learning from participating in the project. The focus on outcomes for pupils was paramount to our thinking and provided the centre piece of our integrated impact approach, which here we call Target Oriented Lesson Study:

Figure x: Integrating evaluation and improvement



The above approach therefore mixes the planning, application, learning and reflection on lesson study with formal evaluation. The key tools used in this process are outlined below:

i) The impact frames

First, teachers were asked to identify between 3 and 6 case pupils in their classes to focus their efforts on for the year. For these pupils, they were asked to describe the *baseline situation* in relation to a specific aspect of their learning, for instance problem solving in mathematics. They then made qualitative statements that described what these pupils said, wrote, what they achieved and how they appeared to feel in relation to learning. Teachers had to avoid using general terms such as 'the pupil is not motivated', instead, they would write, 'pupil x frequently does not complete written tasks' or 'pupil x frequently writes sentences that show incorrect use of adverbs', for instance. We also discouraged the use of jargon terms like 'peer learning', instead using simple and specific language, such as, 'pupil x usually does not ask his partner for help to select an appropriate strategy to solve mathematical problems'. Similarly, we asked for statements about baseline teaching practice. The next step was to ask teachers to imagine that it was the end of the academic year, and to say what the pupil is now doing, saying, writing, feeling and achieving and in relation to this, what they as teachers would now be doing, saying and feeling. These 'impact frames' (IFs) followed a simple format, and an example of a blank one is given in appendix one. We took in some impact frames at the start of the year and gave feedback, asking teachers to amend as necessary. At the end of the year, we asked teachers to colour-code each of their impact statements, to show red: not achieved, yellow: partially achieved or green: fully achieved. These gave us a rich source of information about the types of impact that teachers achieved with their pupils over the course of the year

ii) The impact of lesson study cycles

Each lesson study cycle had a 'focus teacher', i.e. the teacher taking the class. They would typically be helped to plan their lesson with two other colleagues and in some cases these would be facilitated by a trained 'lesson study lead' from another school or an external consultant, employed by the university. The focus teacher would use a lesson study planning sheet devised for the programme and would also refer to their IF. The IF provided a useful overall focus, within which the lesson plan would have a subordinate aims that linked to this. The lesson planning was often a shared focus between the three teachers and there would be intense discussion about the context of the class and in particular about the case pupils. After the second lesson and post lesson discussion, the focus teacher would take detailed notes and complete the focus teacher questionnaire. This included making entries that commented on the observations about each case study pupil, evidence about changes to their learning and the extent to which they saw improvement to learning on the scale below:

No change (or worse than expected)
Improvement to a specific aspect of learning in this subject
Improvement which should impact on students' achievement in this subject
Profound transformation to this student's learning in this subject

Given the gap between the first lesson in the cycle and the second lesson, some teachers have clearly made comments that chart improvement throughout this cycle too, although this may reveal

some lack of clarity about the nature of the evaluation. An examples of this was: increased attendance. In addition, the entry regarding changing of the seating was more about the cause of the improvement in learning rather than the outcome.

iii) Teacher self-efficacy measures

Teachers were asked to self-assess their confidence on nine areas of pedagogy. These nine areas were based on a report by Husbands, C. and J. Pearce (2012) that reviewed a range of literature for the then National College for School leadership[4]. This was converted into a baseline and end of year self-assessment. All teachers taking part in the LS cycles were urged to complete these measures as a way of charting their professional learning and perceived efficacy in pedagogy. The measures served two purposes. First, given the dominant influence of Ofsted[5] on teachers' ideas of what constituted a 'good' or 'outstanding' lesson, we wanted a measure of learning that was informed by a wider review of the literature. This would also avoid a circular logic that teachers whose students achieved high grades were the best teachers, i.e. begging the question what the features are necessary to achieve this (narrowly-defined) goal. The second purpose was to serve as a self-evaluation for teachers, who could then think about the skills and knowledge they may want to focus on over the year. In particular, we wanted teachers to think about the skills they might need in order to achieve the aims they had set out for their case students. The nine areas were:

1) Consideration to pupil voice, i.e., engaging pupils in decisions about the teaching and learning process and responding to their feedback

2) Understanding of the pedagogical process, i.e., understanding the link between subject-matter knowledge, teaching processes and knowledge (and beliefs) about children and their development

3) Clear thinking about longer term learning outcomes as well as short-term goals, i.e. understanding how individual lessons fit to a longer-term sequence of learning outcomes

4) Building on pupils' prior learning and experience, i.e., leading learning from the pupils' current developmental level to a higher level of potential development

5) Scaffolding pupils learning, i.e., supporting developmental changes in learners, including social and emotional support and fostering motivation

6) Using a range of techniques, including whole-class and structured group work, guided learning and individual activity, i.e, planning, organising and implementing a range of effective teaching approaches to advance pupil learning?

7) Developing higher order thinking and meta-cognition, i.e. making good use of dialogue and questioning in order to advance students' understanding of their learning processes

8) Embedding assessment for learning, i.e. using student achievement to inform next steps in instruction and helping students to be self-regulating and working to their own goals rather than by comparing to others

9) Inclusivity, i.e. taking into account the needs of a diverse range of students; encouraging all students to achieve equally highly; and avoiding crude categories and stereotypes

[4] Now re-named the National College For Teaching and Leadership.
https://www.gov.uk/government/organisations/national-college-for-teaching-and-leadership
[5] Office for Standards in Education is the external inspectorate for state schools in England

iv) Interviews with leaders of lesson study

In LCK, we had trained a first cohort (ch1) in the LS process and in how to lead this with a trio of teachers in a neighbouring school. In Ascend, all LS cycle occurred within schools and the trios were put together using a range of criteria, mostly determined by the Headteacher of each school. In order to find out about the organisational level for implementing effective LS cycles we interviews the CH1 LS leaders in LCK and also Headteachers and other senior leaders, as well as participating teachers, in situ, at 3 participating schools, reflecting primary and secondary phases. These interviews were particularly important to determine the factors that would ensure successful implementation beyond the duration of the project. We were also strongly minded to focus on the pedagogical leadership skills required to maximise impact of LS and this covered middle and senior leadership levels.

v) Overall lesson study survey

We were mindful that LS cycles were carried out at different times of the year and that teachers were involved not only as focus teachers but also as co-planners and observers of others' lessons. Therefore we wanted to capture the key professional learning features of LS cycles over the course of the year. These were captured by responses to seven statements on a four point Likert scale and one open ended invitation for comments.

vi) Quantitative Pupil outcomes

This aspect was the most challenging for several reasons. Prior to starting year two of the projects, the government had decided that schools should no longer used long-established and well-understood measures of progress. This became known as 'assessment without levels' (McIntosh, J. 2015). While having laudable aims, this meant that we no longer had a measure of progress that all primary school participants could agree on across the projects. In the case of Ascend, we also had secondary schools, so these would measure progress and attainment differently too. Our approach to this was to say that we would use a nominal scale of measurement, determined by each teacher, which would allow for them to say if a target was achieved for each of their case students. Essentially this meant that for each student, the teacher would assess the baseline, measure progress for a few months, and then project this rate of progress to the end of the year. The teacher would then try to outstrip this projection, setting an aspirational target. As long as the teacher could say if the target was met, and that this was greater than the projection, they state yes or no in the appropriate column of a spreadsheet.

(insert graph here?):

Table x: Summary of data collection tools and response rates

Data collection tool	Level of impact	Ascend project	LCK project	Combined data	Metric used
Written analysis and self-evaluation using 'impact frame' tool	Changes to teacher use of knowledge and skills Changes to pupil learning outcomes	30 carried out impact assessments	23 teachers carried out impact assessments	53 teachers (relating to over 160 target pupils)	Teachers record baseline and impact descriptions of teaching practice
Teacher self-evaluation surveys	Changes to teacher use of knowledge and skills	49 teachers completed baseline and final self-evaluations	23 teachers completed baseline and final self-evaluations	72 teachers	Nine item scale of 1-7 for each (7 highest). Baseline and end of year evaluations
Overall lesson study survey	Teacher professional learning Teachers' reactions: Attitudes and enjoyment	55 teachers completed online survey	23 teachers completed online survey	78 teachers	3 open ended questions on teacher learning and 8 Likert Scale responses (4 point scale Strongly Agree-Strongly Disagree) on reactions to LS process
Focus teacher surveys	Pupil learning outcomes during lesson study cycle	Data from 40 teachers and 120 case study pupils	Data from 26 teachers and 101 case study pupils	66 teachers and 221 pupils	Record of observations about pupils and assessment of extent and type of improvements to learning on 5 point scale
Focus group interviews on leading Lesson Study	Organisational support and change	n/a	6 C1 teachers interviewed in group	n/a	Open responses in group interview on lessons learned about leading LS
Interviews with school leaders and teachers in a sample of schools	Organisational support and change	3 case schools visited	n/a	n/a	Semi-structured interview prompts
Quantitative Changes to pupil progress	Pupil learning outcomes	236 pupils	98 pupils	334 pupils	Achievement of target level (set by teachers at start of year)

The efficacy of the evaluation tools:

Here we will comment primarily on the utility of the methodological approaches and less on the results of the projects themselves; a detailed assessment of the results is available elsewhere (authors, forthcoming).

i) Impact frames

These impact frames immediately focused minds on what teachers were hoping to achieve with their pupils. Interview and survey feedback from teachers in year one, showed the value of this in a) helping to plan the lesson for the LS cycle and b) providing specific aspects for colleagues to look for when observing pupils and c) to gauge the success of teachers approaches over the year with their target pupils. Qualitative analysis of Impact Frameworks was carried out using NVivo (N = 53) on the colour coding given by teachers to each of their statements. Participants largely felt that they had achieved or partially achieved their goals:

	% Achieved	% Partially Achieved	% Not Achieved
Changes to practice	65	30	5
Changes to pupil outcomes	57	36	7

Changes in practice described by participants were analysed using nine categories, based on Husbands and Pearce (2012), together with an additional category for 'other teacher competences'. There were 155 impact statements describing the impact of the lesson study project on the teaching practice of the 53 teachers. 108 (70%) of these statements were categorised as either Embedding Assessment for Learning (23%), Inclusivity (24%) or Scaffolding Pupil Learning (23%). Of the 108 impact statements in these three areas, around two-thirds were Achieved, almost a third were Partially Achieved, whilst only 7 were Not Achieved. The remaining 47 impact statement were more thinly spread across 5 other areas. No impact statements were coded as Clear Thinking about Longer Term Outcomes, nor as Understanding the Pedagogical Process; this is an interesting finding since these areas were still perceived as significant areas of development in the teacher self-evaluations.

There were 468 impact statements describing the impact of the lesson study project on the pupils' experience and learning. 86 (18%) of these statements were categorised as Number. No impact statements were coded as Geometry, Measurement or Statistics (from Programme of Study) while 95 statements (20%) were coded as either Problem-solving in Mathematics or the Use of Spoken Language in Mathematics ('hidden curriculum'). Of the 86 impact statements in Number, around half were Achieved, over a third were Partially Achieved, whilst only 9 were Not Achieved. 58 (12%) of the statements were categorised as Writing, whilst 29 (6%) were shared between Reading and Spoken Language. Of the 58 impact statements in Writing, almost two-thirds were Achieved, almost a third were Partially Achieved, whilst only 5 were Not Achieved. 73 (16%) of the statements were categorised as Engagement. The remaining 127 impact statements (27%) were spread across Collaboration, Confidence, Perseverance and Resilience. Of the 73 impact statements in Engagement, over half were Achieved, over a third were Partially Achieved, whilst only 3 were Not Achieved.

Pedagogy

Changes in practice described by participants were analysed using nine categories, based on Husbands and Pearce (2012), together with an additional category for teacher competences identified which did not easily fall into any of the nine categories.

ii) The impact of lesson study cycles

Table x: Scale of extent of improvement to case pupil learning during a lesson study cycle:

No. of entries about case pupils	Nature and extent of change		
Total	No. of pupil entries	%	Response
301	40	13%	No change (or worse than expected)
	129	43%	Improvement to a specific aspect of learning in this subject
	120	40%	Improvement which should impact on students' achievement in this subject
	12	4%	Profound transformation to this student's learning in this subject

The table above shows, perhaps unsurprisingly, the majority of statements relating to improvements to specific aspects of learning and those that could have a lasting impact on subject knowledge, with only a small percentage relating to profound transformations. The fact that 13% of observations showed no improvement or worsening again should not be too surprising during one single lesson study cycle. An interesting area for development here may be in how teachers can look for key aspects of the pupil's learning that are likely to lead to the biggest changes in their progress over time and to see the LS cycle as an exercise in gaining insight to these aspects. Hence, the nature of these observations is interesting when coded into particular themes (see below).

Table x: coding of entries for types of improvement in learning (add to app?)

No. of entries	%	Coded themes
88	17%	Increase in understanding/usage of knowledge
82	16%	Increase in engagement/participation
72	14%	Positive impact on confidence
46	9%	absent/no clear change in any aspect
34	6%	Increase in communications (pupils vs pupils and pupils vs teachers)
34	6%	Positive impact on independence of pupils
33	6%	Positive impact on concentration level
30	6%	Improvement in quality/quantity of work completed

29	5%	Demonstrating evidence of good use of learning aids (manipulatives, iPads, etc)
21	4%	Demonstrating benefits from peer support
18	3%	Ability to answer questions
17	3%	Increase in resilience
8	2%	Change in seating location (that led to improvement in learning/engagement)
7	1%	Other
5	1%	Improvement in body language
528	100%	Combined coded themes

These results were extremely helpful in our evaluation report to the funders of the projects, whose primary concern was with increasing subject knowledge. The evidence in observations from LS cycles shows considerable success in this area with case students. The concerns for raising confidence and the level of participation of students, often went hand in hand with the first concern and strategies that were used by teachers often reflected this. We can also see the value in having extra eyes in the classroom, where detailed insight was gained about aspects such as students' use of resources, body language, use of questions and increases in communication.

iii) Teacher self-efficacy measures

Across the project we saw high or medium effect sizes for shifts in each of the nine pedagogical areas - * Combine effect size data from Ascend and LCK (Fabian)

The interesting aspect here was to test the internal validity in the use of these nine evidence-informed principles of good pedagogy as a self-evaluation measurement. Therefore, principal component analysis was carried out to operationalise the responses on the 9 variables into a single factor score. Based on the moderate and positive correlations observed between these items (see appendix x), a unique factor was successfully extracted. This measure accounted for by 52.4% of the total variance on these indicators, which in turn yielded meaningful factor loadings for all the items (all over .6). These results suggested that these indicators were tapping a single construct related to high-quality pedagogy; in other words, the higher the scores in all these 9 items, the higher the values in the underlying factor score.

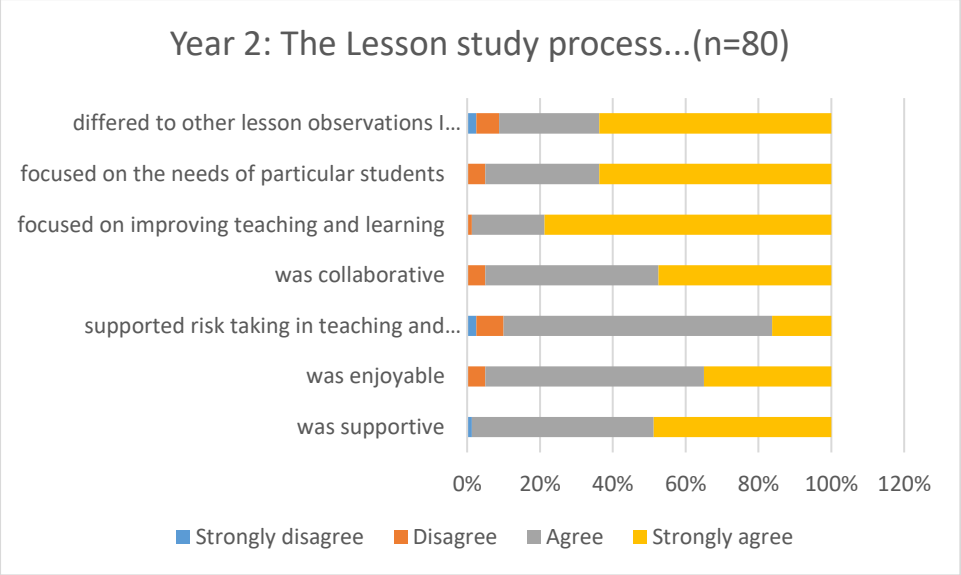
iv) Interviews with leaders of lesson study

Interviews that took place in schools for Ascend revealed concerns about the consistency with which school leaders were supporting the efforts of teachers to work collaboratively in LS cycles over the school year. Key to the success of LS was perceived to be: protected time for teachers to work together and leadership management of LS cycles. In some schools teachers were asked to find time after school or other 'free' periods to discuss the lesson they had planned together and observed. Where LS cycles were more effective, someone in middle or senior leadership helped coordinate who would work with whom, and when to meet up and discuss or plan each phase of the LS process. Protecting teachers from other

staff meetings was key to this. In some cases, this was seen as impractical and school leaders used a faculty structure to designated which faculty would focus on using LS in that academic year, while others would be on a 'normal' cycle. Those on LS cycles could also be temporarily removed from the requirement to be observed as part of a performance management review. This enabled trios to work together in full trust and to take risks, thus focusing on learning from peers, rather than 'proving' they were competent to their seniors. LS leaders in the LCK project reflected on the relational skills that were needed when working with trios outside their school to build trust and also on the value of having a shared theme. These evaluations were useful at the interim stage of the project to fine tune our advice to teachers and also to provide guidance to senior leaders in how to get the most from lesson study.

v) Overall lesson study survey

Figure x: Reflections on professional learning experience throughout the year



The bar chart above shows strong agreement to most statements and little disagreement. Fortunately, most participants felt that the process was collaborative, focused on the needs of their students and improving teaching and learning. Weaker agreement was towards risk taking, although agreement was still fairly high overall. Comments reflected the importance of good relationships and participants felt that risk taking grew once trust was built up in the group. One of the respondents commented on the richness of the learning experience when LS was done well:

"It was very useful to focus on the children's learning when observing lessons rather than the teacher/teaching and to be able to observe how children approach challenges and problem solving. Risk taking was definitely encouraged and it was a safe environment to do so as the lesson was collaboratively planned with all members in the cohort sharing the responsibility."

The focus on pupils was seen as key; professional learning was largely seen as a product of the efforts made to improve pupil learning, this was the key motivation too. By focusing observations on pupils, the teacher also felt less threatened and judged. In the English context, this is particularly important, given the emphasis by external inspectors or senior

leaders on grading and assessing the 'performance' of teacher and teaching quality in lessons in classroom observations.

vi) Quantitative Pupil outcomes

Analysis of both sets of data showed significant gains to pupils' learning, with many achieving the aspirational targets set by their teachers. However, this area was problematic in a number of ways: teachers' self-devised scales used different ranges and different measures, leading to poor reliability and comparison; some teachers appeared to set highly aspirational targets while others were less ambitious or even seemed to reflect a mis-understanding about the nature of the target. In others cases, the teacher's own scale had insufficient range for differentiation of progress to be properly ascertained. Therefore, we are much less confident about this data than the qualitative data obtained from other tools. A wider question remains about what kind of measure would be appropriate, and how to measure these across a whole school, especially when each teacher had varying aims for their case students. The lack of control pupils, matched on key variables is also a problem. Finally, we did not look at the attainment or progress of the non-case students. This raises questions about how the learning of teachers during LS cycles can and should be expected to have a measurable impact on attainment over one academic year for particular students.

Discussion

Strengths and Limitations of our integrated impact model (all)

Strengths	Limitations
Helped us to see qualitative impact of lesson study	Lack of control group means the extent of improvement in teacher self-evaluation is difficult to assess
Could gauge both extent and type of professional learning of teachers	Difficult to have full participation in some surveys, such as focus teacher one
Pedagogical gains were measured and the nine items appear as a single construct	Quantitative gains to pupil progress hampered by lack of clear progress measures or control groups
IF helped frame the LS cycles, the planning, evaluation and impact of them	How to quantify if pupil gains may take longer than one year?/gains are mainly to teacher understanding and skills that may benefit next year's pupils!
Focus on pupil qualitative gains enabled us to evaluate the success of the year and also to make visible the impact of the teachers' efforts	Attrition rates high, especially for Ascend project, where there was less hands-on facilitation by researchers
	Reliance on self-evaluation
	Making sense of the gains to non-target

In summary, we feel we captured well, the logical chain of development from initial teacher reactions to LS involvement through to student level outcomes over the course of a year. The evaluation tools helped us understand the features of LS professional learning that were effective, the organisational aspects that helped create effective lesson study and the strategies and skills that helped improve student learning outcomes. We can also see some modifications that would further enhance our approach. Some key strengths and weaknesses have come from our approaches:

Strengths:

Using our 'impact frame' approach served as an excellent starting point for the whole exercise. It helped teachers to make sense of the whole point of their efforts, i.e. towards improving students' learning.

The teacher self-evaluation tool looks promising as a single construct for pedagogical learning

Interviews with participating staff and leadership help to provide much needed feedback loops on the process.

Having a range of measures helps to capture the range of learning and leadership issues needed for LS cycles to be effective and to lead to pupil learning gains.

Weaknesses:

Our measures still rely heavily on self-evaluated change and improvement. This calls for tests of reliability of these measures as well as validity. Control groups could be useful, in particular, for pupil outcomes and for the teacher self-evaluation too. External funders could then get a sense of what might be unique to LS compared to other forms of PD.

We have not captured the learning of 'non-case study pupils'. While there is some inherent logic that improved professional learning and skills would help all pupils, we have not captured that here. This is partly also due to the lack of a single valid and reliable way of measuring this.

There was also a tension between having 'static' desired outcomes and baseline and the reality of a dynamic and evolving picture. While the impact frame helped immensely to make sense of why teachers were conducting LS cycles, as teachers learned about their case study pupils, they often revised their initial idea of both the baseline and desired impact. This means our impact frame ought to have a separate section for 'non-anticipated' or 'evolving' targets and impact about their pupils and practice. In one of the projects, the use of a learning diary was a really useful additional feature and the sharing of pictures, artefacts and reflections provided further helpful professional learning discussion at our last conference.

Some recommendations on the effective use of an integrated evaluation approach to LS for professional development:

Teachers need critical guidance and feedback on their stated baseline and impact measures. This helps guide observations and avoids vague impact targets that are impossible to evaluate. Impact targets should start with what teachers want to achieve with their pupils, rather than on their own teaching practice, especially since they will not know which strategies they are likely to employ or that are most effective, until they have tried a LS cycle.

Give clear guidance on the kinds of evidence of improved learning that might be recordable during one LS cycle. Observers need to separate out the causes of improvement and the observable indicators of improved learning. Initial debate about this would lead to very high quality PD exercise for teachers too.

Share the results of interim evaluations with both participating staff and school leaders. The former will have a clearer understanding of how to maximise learning from LS cycles and the latter will understand how to facilitate this (and when to keep out of it!).

Explain the evaluation process at the outset and provide timely reminders on how to complete these. This worked best when we incorporated staff development sessions with description of evaluation tools, so that participants did not see these as merely requirements of external evaluators. Where possible, the completion of the evaluations should also take place during the staff meetings, as these were excellent tools for self-reflection and discussion too and enabled high return rates.

Consider the best quantifiable measure of progress for pupils. Is this to do with resilience, confidence or specific aspects to a subject, such as 'problem-solving' in mathematics? If this is a shared theme throughout the project it would be reasonable to see some gains on these features in one school year while overall subject gains may be less detectable in the short term.

We can also see some significant areas for further development and research, including:

- Developing the link between learning from observations of case pupils to whole class

- further refining the teacher pedagogy scale, for instance sub-dividing statements that are loaded with separate items, e.g. number 8 contains aspects of self-regulated learning as well as teachers adapting lessons to feedback on pupils' progress
- the development of tools to measure a range of pupil outcomes
- looking for cross-validity between these tools
- developing tools that measure changes in pedagogical outcomes that do not depend on self-evaluation

Overall, we remain committed to an approach that builds in evaluative thinking to an educational innovation (Earl, L. and H. Timperley 2015), rather than adds evaluation for impact as a 'bolt-on' and urge the use of LS as an innovation that becomes one of a range of tools designed to address the central issue – i.e. improving children's education.

References (to update – David and Sarah)

Dudley, P. (2014) How lesson study works and why it creates excellent teaching and learning. in Dudley, P. (ed.) (2014). *Lesson Study: Professional Learning for Our Time*. Routledge. pp. 1-28

Earl, L. and H. Timperley (2015). Evaluative thinking for successful educational innovation. Education Working Papers. OECD Publishing, Paris. **122**.

Education Endowment Foundation. (2016). *Lesson study.* Available at: https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Round_3_-_Lesson_Study.pdf [Accessed on 08.04.16]

Husbands, C. and J. Pearce (2012). "What makes great pedagogy? Nine claims from research." Research and development network major themes: Theme **1**.

Janis, I. (1972). *Victims of Groupthink.* New York: Houghton Mifflin.

McIntosh, J. (2015). Final report of the Commission on Assessment without Levels, September 2015.

Xu, H. and Pedder, D. (2014) Lesson study: and international review of the research. in Dudley, P. (ed.) (2014). *Lesson Study: Professional Learning for Our Time*. Routledge. pp. 29-58

Appendices:

Appendix x: Calculations for factorial analysis of pedagogical ratings

Pedagogy (PED):

1) Consideration to pupil voice

2) Understanding of the pedagogical process

3) Clear thinking about longer term learning outcomes as well as short-term goals

4) Building on pupils' prior learning and experience

5) Scaffolding pupils learning

6) Using a range of techniques, including whole-class and and structured group work, guided learning and individual activity

7) Developing higher order thinking and meta-cognition

8) Embedding assessment for learning

9) Inclusivity

Table 1. Multiple correlation matrix

	PED1	PED2	PED3	PED4	PED5	PED6	PED7	PED8
PED1	1.0000							
PED2	0.6102	1.0000						
PED3	0.4085	0.5075	1.0000					
PED4	0.5293	0.5432	0.5478	1.0000				
PED5	0.4257	0.4689	0.4486	0.3883	1.0000			
PED6	0.4272	0.4958	0.4789	0.5612	0.3834	1.0000		
PED7	0.5425	0.5384	0.5365	0.5599	0.4672	0.4851	1.0000	
PED8	0.4942	0.4610	0.3901	0.5064	0.4202	0.3447	0.5614	1.0000
PED9	0.3833	0.4091	0.3810	0.4554	0.4067	0.2721	0.3644	0.3952

Factor analysis/correlation Number of obs = 211

Method: principal-component factors Retained factors = 1

Rotation: (unrotated) Number of params = 9

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	**4.71492**	**3.92590**	**0.5239**	**0.5239**
Factor2	0.78903	0.10356	0.0877	0.6115
Factor3	0.68546	0.05944	0.0762	0.6877
Factor4	0.62602	0.04259	0.0696	0.7573
Factor5	0.58343	0.08623	0.0648	0.8221
Factor6	0.49720	0.10591	0.0552	0.8773
Factor7	0.39129	0.01670	0.0435	0.9208
Factor8	0.37458	0.03652	0.0416	0.9624
Factor9	0.33806	.	0.0376	1.0000

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Uniqueness
PED1	0.7469	0.4421
PED2	0.7811	0.3899
PED3	0.7241	0.4757
PED4	0.7897	0.3764
PED5	0.6698	0.5514
PED6	0.6843	0.5317
PED7	0.7854	0.3831
PED8	0.7027	0.5061
PED9	0.6094	0.6286

(blanks represent abs(loading)<.35)