London Schools Excellence Fund

Self-Evaluation Toolkit

Final report

Contact Details
educationprogramme@london.gov.uk

Evaluation Final Report Template

Introduction

The London Schools Excellence Fund (LSEF) is based on the hypothesis that investing in teaching, subject knowledge and subject-specific teaching methods and pedagogy will lead to improved outcomes for pupils in terms of attainment, subject participation and aspiration. The GLA is supporting London schools to continue to be the best in the country, with the best teachers and securing the best results for young Londoners. The evaluation will gather information on the impact of the Fund on teachers, students and the wider system.

This report is designed for you to demonstrate the impact of your project on teachers, pupils and the wider school system and reflect on lessons learnt. It allows you to highlight the strengths and weaknesses of your project methodology and could be used to secure future funding to sustain the project from other sources. All final reports will feed into the programme wide meta-evaluation of the LSEF being undertaken by SQW. Please read in conjunction with Project Oracle's **'Guidance to completing the Evaluation Final Report'**.

Project Oracle: Level 2
Report Submission Deadline: Round 2 - 30 September 2015
Report Submission: Final Report to Rocket Science

Project Name: Physics / Chemistry Foundations
Lead Delivery Organisation: Canterbury Christ Church University
London Schools Excellence Fund Reference: 159
Author of the Self-Evaluation: Mark Hardman
Total LSEF grant funding for project: £117,415
Total Lifetime cost of the project (inc. match funding): £130,285
Actual Project Start Date: 1st March 2014
Actual Project End Date: 31st December 2015

1. Executive Summary

This should be a brief summary of what information is included in the report, the evaluation methods and analysis used and a summary of the key findings from your project evaluation. (maximum 500 words)

Teacher-researcher collaboration enhanced the subject knowledge and pedagogic understanding of a group of teachers who designed a research informed scheme of work (set of lesson plans) for Key Stage 3 physics and chemistry. The process engaged 40 teachers over 11 schools in reading and interpreting research evidence on pupil learning in these subjects. This report argues that focused teacher-researcher collaboration is an effective form of teacher development. This is supported by both qualitative evidence and quantitative pre- and post-testing of the physics group, although the reliability of the latter is limited.

The project also sought to test the impact of these schemes of work on teachers who were not involved in their design, and the pupils they taught. This involved a trial of nine schools who taught the scheme and ten control schools. Only three trial schools (a total of 212 pupils) and three control schools (195 pupils) returned pupil data, and only three teachers from the trial group and three from the control group returned post-test data. As such, no firm conclusions can be deduced from the quantitative trial, although there is a suggestion that the scheme makes no significant difference to outcomes. Qualitative feedback from schools suggests that this is due to the way that teachers adapt schemes of work to the specifics of their classes and settings. It is also possible that the short, closed response, pre- and post-test instruments did not adequately delineate more nuanced impact.

The evaluation considers how greater engagement of senior leaders in schools and a dedicated project officer would have ensured a greater impact of the project and trial, and how professional development should be fostered in an environment conducive to teachers working with researchers and research literature to identify and address their own needs.

The project has attracted considerable attention from teachers and researchers around the world, even before the outcomes have been shared. Although the quantitative evidence is minimal therefore, the weight of qualitative evidence suggests that teacher-researcher collaboration offers a mode of teacher development which is of value to teachers and researchers both within and beyond London schools.

2. Project Description

Much of the detail for this section can be drawn from your Stage 2 funding application. Please note that if you do copy this information from your original application, funding agreement, or interim report, be sure to update it as appropriate (e.g. including tense change).

Provide a full project description (approximately one side of A4), in particular:

- *Why was the project set up? / What need was it seeking to address? (e.g. because teachers lacked confidence in their subject knowledge? Because pupil attainment was lower in this subject area in this borough/cluster/school/than in other boroughs/clusters/schools?).*
- *What were the circumstances into which it was introduced (e.g. existing networks of schools/ expert partner offering a new approach etc.)?*
- *What project activities have been put in place?*
- *Where has the project been delivered geographically?*
- *Who delivered the project?*
- *Who were the target beneficiary groups of the project and why?*

Need for the project

The 2014 Science Programme of Study at Key Stage 3 aims at “building up a body of key foundational knowledge and concepts” (DfE, 2013: p2). Yet with state schools having only 19% of the physics teachers needed (IOP, 2010) the development of conceptual understanding is a significant challenge. Since the 1980s, a growing body of research into ‘conceptual change’ in physics has explored how pupils learn concepts and identified the specific difficulties pupils have, as well as the ineffectiveness of many instructional approaches in overcoming these (Brown & Hammer, 2008).

Evidence base

This project has drawn heavily upon conceptual change research to address this issue. Duit (2009) catalogues over 8000 articles on conceptual change in science, including substantial research on specific aspects of the 2014 Key Stage 3 (KS3) physics curriculum. For example around light (Anderson & Smith, 1983), matter (Babai & Amsterdamer, 2008) and electricity (Clement & Steinberg, 2002), as well as teachers’ concepts in these areas (Küçüközer & Demirci, 2008). Furthermore there is strong evidence around the application of different teaching strategies (Scott, Asoka & Driver, 1991; Smith, Blakeslee & Anderson, 1993). A pertinent example is Gautreau & Novemsky’s (1997) findings that after an initial period spent focusing on conceptual ideas in physics rather than quantitative problem solving, their class performed substantially better than comparison classes on such problems.

Approach

The project was undertaken by a team of three university based teacher-educators (hereafter denoted researchers), who utilised existing contacts with science departments to develop the interventions. We later enlisted the support of a colleague who has expertise in trial data, in order to perform the analysis. There were two main phases of intervention, which throughout this report will be labelled as the *design* phase and *trial* phase.

In the design phase, teachers were engaged in reading groups in which they considered research evidence in relation to specific areas of the physics and chemistry curriculum. In the summer term of 2014, 6 groups (across London) considered 6 aspects of the KS3 Physics programmes of study: Energy; Motion and Forces; Waves; Electricity & Electromagnetism; Matter; Space. In spring and summer 2015, we received ‘scale up’ funding which allowed us to establish similar groups around the KS3 Chemistry programme of study. A further 6 schools (again spread across London) considered: Atoms, Elements and Compounds; Periodic Table; Rocks; Reactions & Energetics; Pure and Impure Substances; The Particulate Nature of Matter. Readings were selected according to the interests of each group and a researcher attended roughly half of these reading group meetings. The physics groups were also supported by e-mail contact with 6 Professors of Physics from Imperial College London, who were available to answer any subject specific questions.

Once they had completed the reading, teachers applied the conceptual change research they had read to the development of schemes of work, in order to embed research evidence in lessons. In doing so the aim was to enhance not just teacher subject knowledge but also pedagogic knowledge around the foundation of concepts which will allow pupils to excel. This provided measurable outcomes in teacher conceptual knowledge (assessed through pre and post-testing) as well as the production of schemes of work in which conceptual change research is embedded. The development of a scheme of work through group study followed a Japanese ‘lesson study’ approach which has been shown to be effective in developing pedagogical understanding (NCSL, 2005).

Despite initially delays in starting the project, we were able to produce the physics scheme of work in time to trial it throughout the 2014-15 academic year. As the chemistry study groups only began in March 2014 (with scale up funding), it was not possible to trial the chemistry scheme of work.

The trial enlisted 10 trial schools and 10 control schools, beyond the schools involved in designing the scheme. These schools were again spread across London, with further schools beyond London also contributing (but not being funded to do so). Pupils’ attainment in physics was measured (using standardised tests), as well as their confidence and enjoyment of physics. This allowed us to test the hypothesis that a research-informed scheme of work, following a conceptual change approach, is able to raise the attainment of pupils beyond ‘conventional’ schemes of work.

Furthermore, we undertook to test the impact that utilising the research informed scheme had on the teachers who taught it, even though they were not engaged in the design. In reality, we were not able to obtain useful data in this area, firstly because of issues in relation to the ‘buy-in’ of schools undertaking the trial and the time taken to complete the pre- and post-tests. Secondly however, we learned that in practice teachers adapt their lessons from the scheme of work such that centralised control of teaching is not possible, and thus the impact of the intervention on each teacher is difficult to measure with the ‘at a distance’ methodologies deployed.

The wider school system will benefit from the availability of a research-informed and conceptually deep scheme of work for KS3 Physics which is accompanied by teachers’ notes. We also uphold the effectiveness of focused reading groups as mode of Continual Professional Development (CPD) which has a high impact and is personalised to the needs of small subject teams.

2.1 Does your project support transition to the new national curriculum? Yes

If **Yes**, what does it address?

The project developed and tested research informed schemes of work for the 2014 Key Stage 3 Physics and Chemistry curricula.

2.2 Please list any materials produced and/or web links and state where the materials can be found. Projects should promote and share resources and include them on the LondonEd website.

The schemes of work produced and further details of the project are available at www.conceptualchange.org.uk

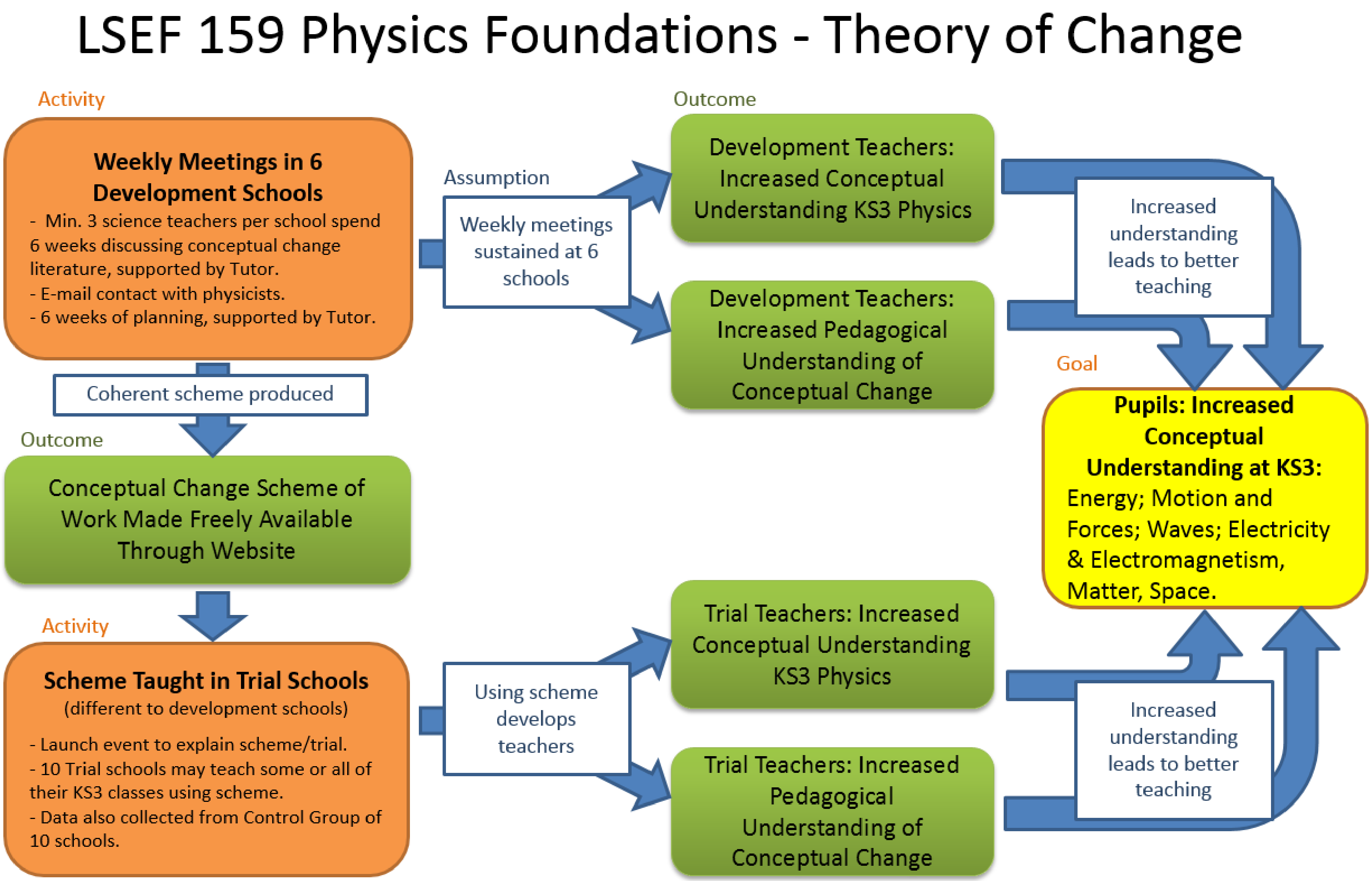
The project was presented as part of a symposium at the European Science Education Research Association conference in Helsinki, Finland in September 2015 (www.esera2015.org). The slides appear on the sharing site http://ow.ly/SriUR

We intend to publish the findings in an academic journal.

3. Theory of Change and Evaluation Methodology

Please attach a copy of your validated Theory of Change and Evaluation Framework.

Throughout the report it would be useful if you make reference to these documents. Where appropriate we would also encourage you to include any assumptions you have made from previous research.


LSEF 159 Physics Foundations - Theory of Change
Activity
Weekly Meetings in 6 Development Schools
- Min. 3 science teachers per school spend 6 weeks discussing conceptual change literature, supported by Tutor.
- E-mail contact with physicists.
- 6 weeks of planning, supported by Tutor.
Assumption
Weekly meetings sustained at 6 schools
Outcome
Development Teachers: Increased Conceptual Understanding KS3 Physics
Development Teachers: Increased Pedagogical Understanding of Conceptual Change
Increased understanding leads to better teaching
Goal
Pupils: Increased Conceptual Understanding at KS3: Energy; Motion and Forces; Waves; Electricity & Electromagnetism, Matter, Space.
Coherent scheme produced
Outcome
Conceptual Change Scheme of Work Made Freely Available Through Website
Activity
Scheme Taught in Trial Schools
(different to development schools)
- Launch event to explain scheme/trial.
- 10 Trial schools may teach some or all of their KS3 classes using scheme.
- Data also collected from Control Group of 10 schools.
Using scheme develops teachers
Trial Teachers: Increased Conceptual Understanding KS3 Physics
Trial Teachers: Increased Pedagogical Understanding of Conceptual Change
Increased understanding leads to better teaching

Evaluation Plan – LSEF159: Physics Foundations

Outputs	Indicators of Outputs	Baseline data collection	Impact data collection
a) Launch meeting during which project is set up and literature introduced.	Consensus achieved on division of scheme and process.	Attendees from each school (online sign up).	Register of attendance. Minutes of meeting. Analysis of churn.
b) Study meetings to engage with conceptual change literature (6 meetings). Supported by tutor & e-mail input from Physicists at Imperial College.	Notes from meetings show progression in conceptual understanding.	Initial conceptual understanding levels and confidence levels. Initial pedagogical understanding. All included in teacher pre-test, taken by all participating teachers.	Registers of attendance and notes from each meeting returned to tutors. Post-test taken by all teachers to assess conceptual understanding, confidence, pedagogical understanding (at end of development phase – July 2014).
c) Planning meeting in which format of scheme is decided	Consensus achieved on division of scheme and process. Standard format of plans & resources agreed upon.	Attendees from each school (online sign up)	Register of attendance. Minutes of meeting. Analysis of churn.
d) Meetings to plan 1/6 of the Scheme of Work. Supported by tutor & e-mail input from Physicists at Imperial College.	Teacher's notes and lesson plans from meetings show application of conceptual understanding to planning. Scheme has clear links between conceptual change and activities.	Initial conceptual understanding levels and confidence levels. Initial pedagogical understanding. All included in teacher pre-test, taken by all participating teachers.	Registers of attendance and notes from each meeting returned to tutors. Post-test taken by all teachers to assess conceptual understanding, confidence, pedagogical understanding (at end of development phase – July 2014).
e) Scheme materials to be made available to trial schools with supporting resources and materials, so that it can be delivered by teachers.	Uptake of scheme by trial schools and collection of baseline data (Sept). Positive evaluation of scheme by physics teachers (outside development group).	Teacher pre-testing of conceptual understanding will be linked to teacher biographical and performance data (inc. years of experience and specialism). Pupil biographical data (including KS2 attainment) and school and class context data will be collected by trial schools. Pupil attainment data in conceptual knowledge test will be linked to biographical data and teacher data.	Teacher post-testing of conceptual understanding will be linked to teacher biographical and performance data (inc. years of experience and specialism). Pupil biographical data (including KS2 attainment) and school and class context data will be collected by trial schools. Pupil attainment data in conceptual knowledge test, will be linked to biographical data and teacher data. Evaluation of scheme by teachers and pupils in trial.

Teacher Outcomes	**Indicators of Outcomes**	**Baseline data collection**	**Impact data collection**
Increased subject knowledge and greater awareness conceptual change approaches (all of development group)	Increased teacher scores in subject knowledge/ teaching approach tests Tests to be taken by all teachers involved in the development phase- is this the same as the trial group? No, it is the development group: 6 schools writing the scheme. What test is this? Who designed/reviewed? We used questions taken from the literature on teacher Pedagogical Content Knowledge (PCK) supported by peer reviewed publications. Designed by project lead and reviewed by another Senior Lecturer in Education	Scores collected for individual teachers from pre intervention subject knowledge/ teaching method tests, linked to biographical data. Before start of project: April 2014	Scores collected for individual teachers from post development phase intervention subject knowledge/ teaching method tests, linked to biographical data. Qualitative data from session notes. July/August 2014
Increased teacher confidence in KS3 physics (development group)	Increased teacher scores in confidence surveys What confidence survey are you using? An existing tool or self-developed? If self-developed, please specify who designed and reviewed. Teacher confidence surveys use format provided by GLA Survey to be completed by all teachers involved in the intervention	Scores collected for individual teachers from pre intervention confidence surveys Before start of project: April 2014	Scores collected for individual teachers from confidence surveys after development phase. Qualitative data from minutes of meetings (tutors attend fortnightly). July/August 2014
Increased subject knowledge, greater awareness of conceptual change approaches (trial group)	Increased teacher scores in subject knowledge/ teaching approach tests Tests to be taken by all teachers involved in the trial and control groups. What test is this? Who designed/reviewed? This is the same test as given to the	Scores collected for individual teachers from pre teaching subject knowledge/ teaching method tests, linked to biographical data. Before start of trial: Sept 2014	Scores collected for individual teachers from post teaching subject knowledge/ teaching method tests, linked to biographical data. Qualitative data from teacher survey – sample of trial and control group. What survey is this? What sample size will you use? How will you choose sample?

	development group (see above). Therefore it has been through a further stage of testing also.		July 2015 (Analysis by Nov 2015)
Increased teacher confidence in KS3 physics (trial group)	Effect size on confidence relative to control group. Note: negative effect size may be linked to breaking down of misconceptions. Thus, linked to qualitative survey data. Who is in the teacher control group? How have they been matched? Schools in the control group are self-selecting so teachers will be matched retrospectively across a range of dimensions: experience, subject specialism, school type (Fischer Family Trust data). We will also be matching pupils and teacher data so can match across pupil types (see below)	Scores collected for individual teachers (intervention and control group?- Yes) from pre teaching confidence surveys Before start of trial: Sept 2014	Scores collected for individual teachers from confidence surveys after teaching the scheme (intervention and control group?). Qualitative data from teacher survey – sample of trial and control group. Sample size? Sampling method? Change of plan: give survey to all teachers in trial and control group (will just require more analysis) July 2015 (Analysis by Nov 2015)
Pupil Outcomes	**Indicators of Outcomes**	**Baseline data collection**	**Impact data collection**
Increased conceptual understanding of KS3 physics for all of trial group being taught with scheme.	Significant effect size in assessment above control group. All pupils in control group and trial group to sit test. Who is in the pupil control group? How have they been matched? Schools in the control group are self-selecting so pupils will be matched retrospectively across a range of dimensions (see next box). Pupils will also be matched to teachers of significance of teacher 'type' can be explored.	Pre-test of conceptual understanding of all pupils. Linked to data on KS2 attainment, LAC, FSM, FSM in 6 years, disadvantaged, EAL, gender, ethnicity, SEN. Before start of teaching physics in Sept 2014 What test is this? Who designed/reviewed? Test questions are taken from TIMSS database 2011 & 2009 – questions used for international comparison of science understanding.	Post-test of conceptual understanding of all pupils. Linked to data on KS2 attainment, LAC, FSM, FSM in 6 years, disadvantaged, EAL, gender, ethnicity, SEN. Churn analysis. Data collected July 2015 What test is this? Who designed/reviewed? Test questions are taken from TIMSS database 2011 & 2009 – questions used for international comparison of science understanding.

Pupil engagement with KS3 physics increased for trial group (on average).	Survey to evaluate scheme of work in trial and control groups.	Survey questions (incorporated in test) to assess engagement with physics Before start of teaching physics in Sept 2014 Survey is taken from Kind, Jones and Barmaby (2007)	Survey questions (incorporated in test) to assess engagement with physics. Sample of control and trial group to complete extended evaluation of scheme. Data collected July 2015 Sample size? Sampling method? 20% of cohort in trial and control group: determined by biographical dimensions in pre-test to ensure representative spread of KS2 attainment, LAC, FSM, FSM in 6 years, disadvantaged, EAL, gender, ethnicity, SEN
School System / 'Culture Change' Outcomes	**Indicators of Outcomes**	**Baseline data collection**	**Impact data collection**
Use of new resources by teachers/ schools **outside the intervention group**	Uptake of new resources developed by LSEF programmes by non LSEF teachers/ schools	Planned new resources to be developed by LSEF programmes Avenues of dissemination/ promotion Dissemination for Sept 2014	Number of resources downloaded from websites (by different schools) Number of resources taken from training sessions/ conferences (by different schools) User feedback on quality of resources through online survey Data analysed after July 2015
Programme activities/ model is embedded in department/ schools planning beyond the intervention group (in other subjects within intervention schools and in other schools.)	Inclusion of programme activities/ model in development plans	Development plan pre roll-out of intervention Commitment/ sign up by school to specific criteria pre intervention	Part of department/ school/development plan Number of teachers following development plan/ due to roll out changes Commitment/sign up by school to specific criteria as part of project e.g. release of staff for 3 days of meetings and for study group meetings.

3.1 Please list **all** outcomes from your evaluation framework in Table 1. If you have made any changes to your intended outcomes after your Theory of Change was validated please include revised outcomes and the reason for change.

Table 1a- Outcomes for Teachers in Design Phase (for both Physics and Chemistry)

Description	Original Target Outcomes	Revised Target Outcomes	Reason for change
Teacher Outcome 1a	Increased subject knowledge and greater awareness conceptual change approaches		
Teacher Outcome 2a	Increased teacher confidence in KS3 physics	Also increased teacher confidence in KS3 chemistry	Additional group of teachers considering chemistry

Table 1b- Outcomes for Teachers, Pupils and the Wider System in Trial Phase

Description	Original Target Outcomes	Revised Target Outcomes	Reason for change
Teacher Outcome 1b	Increased subject knowledge and greater awareness conceptual change approaches		
Teacher Outcome 2b	Increased teacher confidence in KS3 physics		
Pupil outcome 1	Increased conceptual understanding of KS3 physics for all of trial group being taught with scheme.		
Pupil outcome 2	Pupil engagement with KS3 physics increased for trial group (on average).		
Wider system outcome 1	Use of new resources by teachers/ schools outside the intervention group		
Wider system outcome 2	Programme activities/ model is embedded in department/ schools planning beyond the intervention group (in other subjects within intervention schools and in other schools.)		

3.2 Did you make any changes to your project's activities after your Theory of Change was validated? Yes

If **Yes**, what were these changes (e.g. took on additional activities?)

We utilised scale-up funding to undertake an additional design phase looking at KS3 Chemistry with 6 further schools.

3.3 Did you change your curriculum subject/s focus or key stage? Yes

If **Yes**, please explain what changes you made, why, and provide some commentary on how they affected delivery.

As above, we added chemistry in addition to physics. This allowed us to engage with a greater number of schools and teachers and also cover another part of the KS3 science curriculum. This makes the final schemes more useful to schools (although biology is still not covered within this project).

The chemistry design group ran in parallel to the physics trial group. This means we were not able to undertake a trial of the chemistry scheme, but it did mean we were able to apply our learning from the physics design phase to the chemistry group.

3.4 Did you evaluate your project in the way you had originally planned to, as reflected in your validated evaluation plan?

We evaluated it as planned. The additional chemistry design phase was evaluated in the same way as the physics design phase, in order to aid comparison. However, the questions in the pre- and post-test obviously reflected a chemistry rather than physics subject knowledge and pedagogy.

4. Evaluation Methodological Limitations

4.1 What are the main methodological limitations, if any, of your evaluation?

This can include data limitations or difficulty in identifying a comparison group. In order to get a realistic idea of the strength of your evaluation, and identify possible improvements, it is essential that you reflect on the strengths and weaknesses of your evaluation.
You should address limitations of the evaluation only, not the project itself - *Every evaluation has limitations, so please be honest. This could include limitations relating to:*

- *The kinds of data you could/ could not collect (and the response rate for surveys)*
- *The size of the sample/ group you are evaluating*
- *The extent to which you felt able to assess the impact of activity on beneficiaries (what changes in attitudes/behaviours/attainment were caused by the intervention and what has been caused by other factors)*
- *Also include mitigating actions for methodological limitations where possible – e.g. alternative approaches or solutions and also how these limitations will affect the evaluation of the project (particularly pupil and teachers outcomes).*

4.1a Design Phase

Matching pre- and post- data
In the physics design phase we deployed a coding system in order to ensure anonymity and encourage teachers to take the pre- and post- tests. This involved the participants using

their mother's maiden name and father's date of birth. Nevertheless, in the physics group we only had three matched responses, out of 21 pre-tests and 7 post-tests. To address this, in the chemistry design phase we asked participants to simply include their names, but removed these during analysis to maintain anonymity. Overall, we therefore have a smaller number of matched pre-post tests than hoped, and have had to use average values, which provide a less nuanced view of impact on teachers.

Balancing returns vs rigour in teacher pre-tests

Aware that teachers are extremely busy, we needed to deploy a pre- and post-test that did not take a considerable time to complete, as this would reduce return rate. After including a survey of confidence measures (using the GLA format to allow comparison against projects) we included just 6 questions in the pre-test and post-test. The pre- and post-tests took teachers approximately an hour each and were completed online. Whilst we felt this was about right in terms of teacher time, we were only able to ask one subject knowledge question per aspect of the curriculum (as well as confidence levels in those areas).

The GLA questionnaire on teacher confidence assesses different aspects of teacher practice across 16 questions. However, there is no triangulation across questions (for example by asking a similar question 'in reverse'), as such this is a limitation. We then added additional questions to assess confidence in each of the pertinent subject areas (e.g. forces or space physics), and in teacher confidence in uncovering misconceptions and using practical work. Again, there is no triangulation across these questions however.

In order to assess subject and pedagogical understanding, 6 further questions followed. Part A of each of these questions in the pre- and post- tests was taken from internationally recognised databases used to assess undergraduates, and we found one question for each of the 6 areas of physics, and again for the 6 areas of chemistry. Following these questions, we asked teachers to consider what difficulties pupils may have (part B of each question) and what strategies they would deploy for addressing those difficulties (part C of each question). We then used a scale to assess their pedagogical understanding (as developed by Riese, Vogelsang & Reinhold (2011)). Whilst we are therefore confident in the approach to each question, there was no opportunity for triangulating teacher responses across questions on the same subject matter (e.g. on forces). Thus, our overall measures of assessing subject knowledge and pedagogical understanding is limited to these 6 questions

Buy in of School Leadership

Because the schools were recruited through the contacts we already had in subject departments, and because they were recruited mid-year, we found that senior leaders were supportive only insofar as the projects activities did not interfere with school level priorities. This is despite us providing funding for cover. This meant that in many of the departments teachers conducted the activities in their own time, with only some being compensated for this (by payment or cover of lessons). As well as two schools dropping out of the project and the need for the researchers to therefore design a small proportion of the scheme themselves, this also had an effect on the response rate from teachers to pre- and post-tests. Teachers often prioritised the activities around engaging with the research and developing the scheme of work, over engaging with the project measure. Although understandable, this affected the return rate of teacher data.

4.1b Trial Phase

Matching pupils and schools in the trial

The schools involved in the trial were recruited through the existing contact networks of the researcher involved. They were thus self-selecting and chose whether they wished to be part of the trial or control group. Therefore, cohort and school effects had an influence, with some schools having lower scores in certain measures (see Appendix 1 for analysis of pupil data). Biographical data was not provided by all of the control schools, which meant that analysis by pupil sub-group was not possible due to the sample size once this had been accounted for. We shall endeavour to obtain this data in order to conduct further analysis.

Engagement of trial/control schools and sample size

Because schools were recruited through existing networks, and often at the departmental level, the return rate on trial and control data is lower than hoped. We received data from three schools trialling the scheme and 3 control schools, out of a possible 10 in each group.

In parallel to the above reflection on buy-in of school leadership in the design phase, we found that those teachers engaging in the trial/control were often doing so in their own time, without the full support that senior leaders could have provided (again, despite funding we provided to the schools). This meant we received pre-test data from two schools who were not then able to provide post-test data, as well as those schools from which we did not receive any data (and subsequently did not provide funding to).

4.2 Are you planning to continue with the project, once this round of funding finishes?

Yes, we are investigating funding streams in order to undertake a design phase for the biology part of the KS3 curriculum.

If **yes**, will you (and how will you) evaluate impact going forward?

We will use the same methods as in the physics and chemistry design phases (with some refinements).

5. Project Costs and Funding

5.1 Please fill in Table 2 and Table 3 below:

Table 2 - Project Income

	Original[1] Budget	Additional Funding	Revised Budget [Original + any Additional Funding]	Actual Spend	Variance [Revised budget – Actual]
Total LSEF Funding	£73,415	£44,000	£117,415	£116,071	£1,343.67
Other Public Funding					
Other Private Funding					
In-kind support (e.g. by schools)	£12,870		£12,870	£12,870	0
Total Project Funding	£86,285		£130,285	£128,941	£1,343.67

Please note that the Actual Spend is the forecast for 31st December 2015
The in-kind estimates were for the provision of teacher time in the original budget. We suspect individual teachers contributed more of their time but have no way of measuring this.

Table 3 - Project Expenditure

	Original Budget	Additional Funding	Revised Budget [Original + any Additional Funding]	Actual Spend	Variance Revised budget – Actual]
Direct Staff Costs (salaries/on costs)	£13,870	£18,800	£32,670	£32,670	£0
Direct delivery costs e.g. consultants/HE (specify)					
Management and Administration Costs	£20,570	£1,500	£22,070	£22,070	£0
Training Costs	£23,970	£17,700	£41,670	£40,326	£1,344
Participant Costs (e.g. Expenses for travelling to venues, etc.)					
Publicity and Marketing Costs	£2,100	£7,800	£9,900	£9,900	£0
Teacher Supply / Cover Costs					
Other Participant Costs					
Evaluation Costs	£25,775	£6,000	£31,775	£31,775	£0
Others as Required – Please detail in full					
Total Costs	£86,285	£51,800	£138,085	£136,741	£1,344

5.2 Please provide a commentary on Project Expenditure
This section should include:

- *commentary on the spend profile*
- *budget changes that have occurred, including the rationale for any changes*

[1] Please refer to the budget in your grant agreement

(Maximum 300 words)

Whilst the variance appears small within this analysis, this is because of two recalibrations of the original budget through the project lifetime. There was an initial underspend of £6,691 within the physics design phase of the project, caused by schools dropping out and reducing a two day workshop to one day. This was accommodated when additional fund was granted for the chemistry design phase. At the end of the project, there was a projected total underspend of £11,607. We agreed a reallocation of £7,800 towards further dissemination activities.

6. Project Outputs

Please use the following table to report against agreed output indicators, these should be the same outputs that were agreed in schedule 3 of your Funding Agreement and those that were outlined in your evaluation framework.

Table 4 – Outputs

Description	Original Target Outputs	Revised Target Outputs *[Original + any Additional Funding/GLA agreed reduction]*	Actual Outputs	Variance [Revised Target - Actual]
No. of schools	6	12	11	1
No. of teachers	24	48	40	8
No. of pupils taught by design phase teachers	2080 (estimated)	4189 (estimated)		

7. Key Beneficiary Data

Please use this section to provide a breakdown of teacher and pupil sub-groups involved in your project.

7.1 Teacher Sub-Groups (teachers directly benefitting counted once during the project)

Please provide your definition for number of benefitting teachers and when this was collected below *(maximum 100 words).*

Table 5a – Teachers benefitting from the programme (Physics Design Group)

	No. teachers	% NQTs *(in their 1st year of teaching when they became involved)*	% Teaching 2 – 3 yrs *(in their 2nd and 3rd years of teaching when they became involved)*	% Teaching 4 yrs + *(teaching over 4 years when they became involved)*	% Primary (KS1 & 2)	% Secondary (KS3 - 5)
Project Total	21	19%	38%	29%	0%	100%
School 1	3	0%	67%	67%	0%	100%
School 2	8	25%	38%	38%	0%	100%
School 3	3	0%	67%	67%	0%	100%
School 4	3	33%	67%	67%	0%	100%
School 5	4	25%	25%	25%	0%	100%

Table 5b – Teachers benefitting from the programme (Chemistry Design Group)

	No. teachers	% NQTs *(in their 1st year of teaching when they became involved)*	% Teaching 2 – 3 yrs *(in their 2nd and 3rd years of teaching when they became involved)*	% Teaching 4 yrs + *(teaching over 4 years when they became involved)*	% Primary (KS1 & 2)	% Secondary (KS3 - 5)
Project Total	21	24%	62%	14%	0%	100%
School 1	3	0%	100%	0%	0%	100%
School 2	4	25%	50%	25%	0%	100%
School 3	3	67%	33%	0%	0%	100%
School 4	4	25%	50%	25%	0%	100%
School 5	4	25%	50%	25%	0%	100%
School 6	3	0%	100%	0%	0%	100%

Table 5c – Teachers benefitting from the programme (Combined Design Groups)

	No. teachers	% NQTs	% Teaching 2 – 3 yrs	% Teaching 4 yrs +	% Primary (KS1 & 2)	% Secondary (KS3 - 5)
Project Total	42	24%	62%	14%	0%	100%

Table 5d – Teachers Completing Physic Pre-test for Trial Group

	No. teachers	% NQTs	% Teaching 2 – 3 yrs	% Teaching 4 yrs +	% Primary (KS1 & 2)	% Secondary (KS3 - 5)
Project Total	21	14%	33%	54%	0%	100%

7.1.2 Please provide written commentary on teacher sub-groups e.g. how this compares to the wider school context or benchmark *(maximum 250 words)*

The teachers engaged within the design phases are broadly representative of the departments worked with, although we did not undertake a full analysis of the latter.

Table 5d reports the constitution of the teachers who provided pre-test data in both trial and control schools. This should be seen as indicative of the teachers involved, but does not constitute a reliable analysis.

7.2 Pupil Sub-Groups (these should be pupils who directly benefit from teachers trained)

Please provide your definition for number of benefitting pupils and when this data was collected below *(maximum 100 words)*

We did not undertake to analyse the specific pupil groups taught by the teachers engaged through the intervention, which was focused upon teacher knowledge and pedagogical understanding.

In the trial, only three schools returned data, of the 9 engaged in teaching the scheme. We therefore report this below, as it forms the trial group sample. However, **this should not be taken as a representative sample of the group impacted in the trial**.

Tables 6-8 – Pupil Sub-Groups benefitting from the programme

	No. pupils	**% LAC**	**% FSM**	**% FSM last 6 yrs**	**% EAL**	**% SEN[1]**
Project Total	*212*	*0%*	*21%*	*48%*	*29%*	*9%*
School 1	*68*	*0%*	*21%*	*38%*	*27%*	*9%*
School 2	*104*	*0%*	*18%*	*51%*	*23%*	*10%*
School 3	*42*	*0%*	*26%*	*57%*	*45%*	*10%*

[1]Following correspondence with the GLA, we defined SEN as those at School Action Plus or Statement level. This more usefully relates to the new system for SEND in which School Action is not classified (pupils instead have an individual plan).
2We defined Low attaining as those achieving a Level 3 at KS2, Middle attaining as those who achieved a Level 4, and Higher attaining as those who achieved a level 5. This was taken from the Science teacher grade where possible, and the average of literacy and numeracy scores otherwise.

	No. Male pupils	**No. Female pupils**	**% Lower attaining**	**% Middle attaining**	**% Higher attaining**
Project Total	*34*	*178*	*12%*	*45%*	*43%*
School 1	*34*	*32*	*33%*	*40%*	*26%*
School 2	*0*	*104*	*3%*	*46%*	*51%*
School 3	*0*	*42*	*12%*	*50%*	*38%*

	% Asian Indian	**% Asian Pakistani**	**% Asian Bangladeshi**	**% Asian Any Other background**	**% Black Caribbean**	**% Black African**	**% Black Any Other Background**	**% Mixed White & Black Caribbean**	**% Mixed White & Black African**	**% Mixed White & Asian**	**% Mixed Any Other Background**	**% Chinese**	**% Any other ethnic group**
Project Total	*4*	*3*	*6*	*7*	*7*	*26*	*10*	*0*	*0*	*0*	*3*	*1*	*2*
School 1	*12*	*6*	*11*	*18*	*2*	*5*	*8*	*0*	*2*	*0*	*2*	*0*	*3*
School 2	*0*	*3*	*3*	*2*	*11*	*47*	*16*	*1*	*0*	*0*	*2*	*2*	*1*
School 3	*0*	*0*	*5*	*0*	*8*	*10*	*0*	*0*	*0*	*3*	*10*	*0*	*5*

	% White British	% White Irish	% White Traveller of Irish heritage	% White Gypsy/Roma	% White Any Other Background
Project Total	*13*	*1*	*0*	*0*	*14*
School 1	*9*	*0*	*0*	*0*	*24*
School 2	*8*	*0*	*0*	*0*	*4*
School 3	*31*	*5*	*0*	*0*	*23*

7.2.1 Please provide a written commentary on your pupil data e.g. a comparison between the targeted groups and school level data, borough average and London average *(maximum 500 words)*

Useful links: London Data Store, DfE Schools Performance, DfE statistical releases

8. Project Impact

You should reflect on the project's performance and impact and use **qualitative and quantitative** data to illustrate this.

- Please complete the tables below before providing a narrative explanation of the impact of your project.
- Please state how you have measured your outcomes (e.g. surveys) and if you are using scales please include details.
- Please add graphical analysis (e.g. bar charts) to further demonstrate project impact on each teachers, pupils, wider system outcomes etc. If you use graphs, please ensure that all charts are explained and have clear labels for the axes (numeric data or percentages, for example) and legends for the data.

Please add columns to the tables if necessary but do not remove any. ***N.B. If your project is collecting data at more than two points and may want to add additional data collection points.***

8.1a Teacher Outcomes – Design Groups

Date teacher intervention started: 28th March 2014 (Physics); 6th March 2015 (Chemistry)

Table 9a – Teacher Outcomes: teachers benefitting from the project

The 1st Return will either be your baseline data collected before the start of your project, or may be historical trend data for the intervention group. Please specify what the data relates to.

Target Outcome	Research method/ data collection	Sample characteristics	Metric used	1st Return and date of collection	2nd Return and date of collection
Increased subject knowledge: physics	Online survey: 6 questions taken from various research sources, designed to test conceptual understanding	18 of the 21 teachers responded to the pre-test and 13 to the post-test. Survey was anonymous but biographical information suggests representative sample.	*Mean score out of 12: two marks in each of the 6 areas of KS3 physics (percentage)*	*Group mean (percentage)=* ***6.11*** *(51%)* *Std. dev. =2.02* *n=18* *(Collected April/May 2014)*	*Group mean (percentage)=* ***8.41*** *(70%)* *Std. dev. =2.35* *n=12* *(Collected July/Sept 2014)*
Increased pedagogical skill: physics	Online survey: 6 questions taken from various research sources, designed to test pedagogic response. Follows Riese, Vigelsang & Reinhold (2011)	18 of the 21 teachers responded to the pre-test and 13 to the post-test. Survey was anonymous but biographical information suggests representative sample.	*Mean score out of 30: five marks in each of the 6 areas of KS3 physics (percentage)*	*Group mean (percentage)=* ***5.67*** *(18%)* *Std. dev. =1.91* *n=18* *(Collected April/May 2014)*	*Group mean (percentage)=* ***12.08*** *(40%)* *Std. dev. =2.78* *n=12* *(Collected July/Sept 2014)*

Increased confidence and efficacy: general teaching	Online survey: using standard survey developed by GLA	18 of the 21 teachers responded to the pre-test and 13 to the post-test. Survey was anonymous but biographical information suggests representative sample.	*Mean score based on a 1-9 scale of 'How much can you..'(1 – nothing, 3 – very little, 5 some influence, 7 – Quite a bit, 9 – A great deal)*	*Group mean = **6.44*** *Std. dev. =1.21* *n=18* *(Collected April/May 2014)*	*Group mean = **6.70*** *Std. dev. =1.39* *n=12* *(Collected July/Sept 2014)*
Increased confidence and efficacy: physics	Online survey: adapted from survey developed by GLA. 6 questions on each area of KS3 physics, one on practical work and one on uncovering misconceptions.	18 of the 21 teachers responded to the pre-test and 13 to the post-test. Survey was anonymous but biographical information suggests representative sample.	*Mean score based on a 1-9 scale of 'How much can you..'(1 – nothing, 3 – very little, 5 some influence, 7 – Quite a bit, 9 – A great deal)*	*Group mean = **6.01*** *Std. dev. =1.42* *n=18* *(Collected April/May 2014)*	*Group mean = **6.52*** *Std. dev. =1.27* *n=12* *(Collected July/Sept 2014)*
Number of strategies	Online survey question, asking for strategies that could promote conceptual change	18 of the 21 teachers responded to the pre-test and 13 to the post-test. Survey was anonymous but biographical information suggests representative sample.	*Mean number of strategies listed in open question*	*Group mean = **2.56*** *Std. dev. =1.58* *n=18* *(Collected April/May 2014)*	*Group mean = **1.91*** *Std. dev. =0.94* *n=12* *(Collected July/Sept 2014)*

Represented Graphically*:*

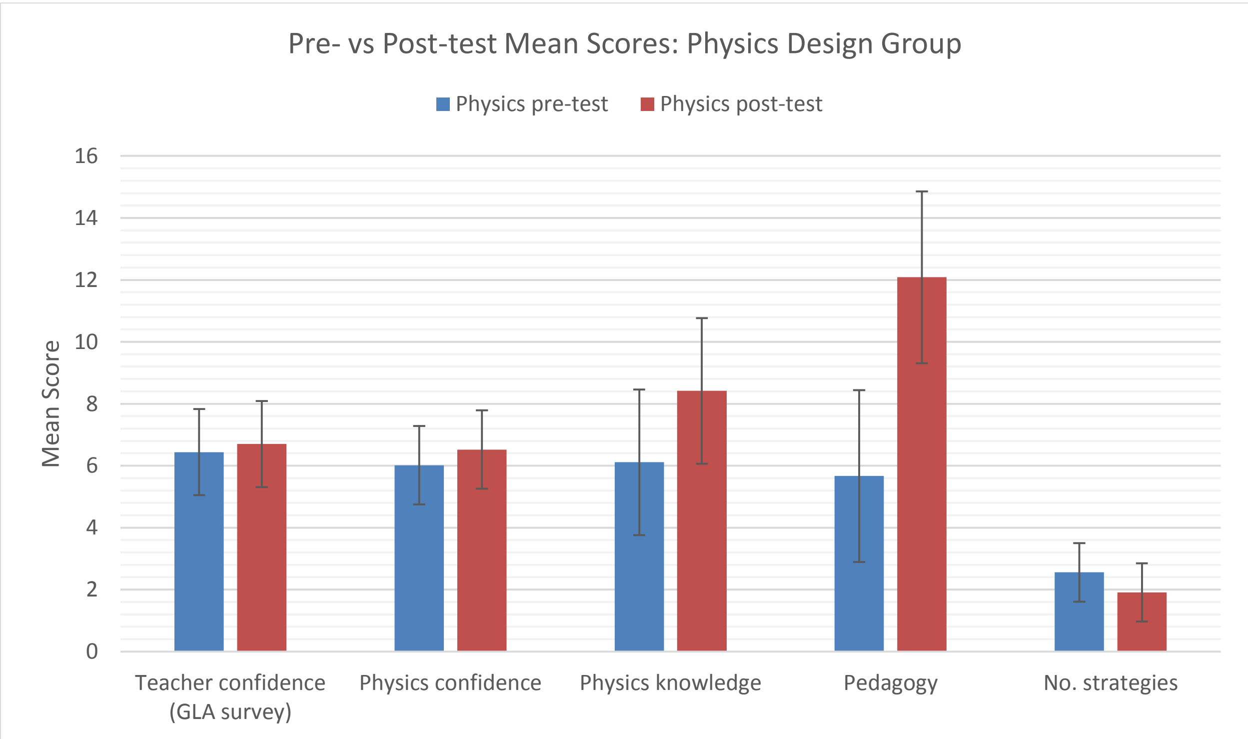


Table 10 – Comparison data outcomes for Teachers *[if available]*

There was no comparison group of teachers within the design phase of the project

Target Outcome	Research method/ data collection	Sample characteristics	Metric used	1st Return and date of collection	2nd Return and date of collection
e.g. Increased Teacher confidence	*e.g. E-survey*	*e.g. 100 respondents from a total of 200 invites.* *The profile of respondents was broadly representative of the population as a whole.*	*e.g. Mean score based on a 1-5 scale (1 – very confident, 2 – quite confident, 3 neither confident nor unconfident, 4 - quite unconfident, 5 – very unconfident)*	*e.g. Mean score*	*e.g. Mean score*

8.1.1a Please provide information (for both the intervention group and comparison group where you have one) on:

- Sample size, sampling method, and whether the sample was representative or not
- Commentary on teacher impact (please also refer to table 5 re impact on different groups of teachers)
- Qualitative data to support quantitative evidence.
- Projects can also provide additional appendices where appropriate.

(Minimum 500 words)

Chemistry Design Group

At the time of writing this report the post-test data for the chemistry group has not been obtained. This is because the development of the scheme is still taking place and as such it would be premature to assess the impact of this process currently. For this reason, the above provides the data for the physics group, who undertook reading groups and design of the scheme of work during spring and summer 2014.

Samples

18 teachers undertook an online pre-test. These tests were anonymised in order to support teachers in reporting their confidence openly and not feeling embarrassment if they could not answer the physics knowledge questions. However, we are able to use biographical data to consider the samples in the pre-and post-test with respect to the overall group:

Teachers Completing Physics Pre-test and Post-test vs Physics Design Group

	No. teachers	% NQTs	% Teaching 2 – 3 yrs	% Teaching 4 yrs +	% Male Female (withheld)	% Physics degree
Project Total	21	19%	38%	29%	19% 76% (0%)	19%
Pre-test sample	18	11%	33%	56%	22% 61% 17%	11%
Post-test sample	12	8%	17%	75%	17% 83% (0%)	17%

The samples can be described as broadly representative of the group, whilst noting that the post-test was completed by a higher proportion of more experienced teachers than the overall group.

The original intention was to provide matched analysis of pre- and post-tests using a code which each teacher developed themselves (first two letters of mother's maiden name and father date of birth). To our surprise, only two of the pre- and post-test codes could be matched in this way, and we cannot account for this. Given that 12 of 21 teachers completed the post-test we would anticipate being able to match more of these to the post-test. Because of this issue, the analysis has had to rely upon mean values from the pre- and post-test, and this is likely to be more sensitive to the differences in pre- and post-test instruments: the differences in the tests.

<u>Commentary on Teacher Impact</u>

Teacher Knowledge

Mean physics knowledge increased by 2.31 (19%), however this is only in line with the standard deviation on this measure so no strong conclusions should be drawn from this result.

Conversationally, teachers reported that their understanding of the particular area of physics they had read about increased and that this had a knock-on effect in other areas. Each area of physics accounts for only 2 marks (17%) in the tests. We are not able to match the outcomes to the areas of study, which would have allowed us to investigate this further.

Teacher Pedagogy & Strategies

The increase in the score for teacher pedagogy was the most notable impact within the design phase of the project. The mean score increased from 5.67 (19%) to 12.08 (40%). This increase is just over twice the standard deviation, which is often considered the threshold of statistical significance. However, without a control group we cannot evaluate the difference between the pre- and post-test. For example, it may have been that the post-test prompted a greater score.

The measure of teacher pedagogy was made up by five marks for each of the 6 subject areas of KS3 physics, so a total of 30 marks. The first three marks for each question were around the number and range difficulties that a teacher could predict a pupil having. The other two marks per question, were about the possible strategies that a teacher had to address pupils misconceptions: a mark of zero was given for the teacher simply explaining/showing the pupils an issue, a mark of one for a pupil led activity and a mark of two if the activity directly challenged the misconception. This follows Riese, Vigelsang & Reinhold's (2011) classification of declarative and procedural Pedagogic Content Knowledge (PCK). The questions asked under each topic were matched, so as to be conceptually the same (e.g. a ball falling from a plane in the pre-test vs a cannon in the post-test), which should have increased the reliability of the score, as it pertained to pupil difficulties in similar scenarios. However, it is still possible that the composition of the test led accounts for some, if not all, of the difference in score.

The measure of the number of strategies a teacher listed under an open question was intended to consider the range of approaches they had to conceptual change. The data show a marginal decrease in this, but this is less than half a standard deviation and is therefore insignificant. However, the question said 'given the above, list the strategies…' which was intended to relate to the whole test, but several teachers seemingly interpreted this as relating to the final topic question only. Furthermore, in the post-test the final question was about space, which is of course difficult to suggest practical pupil activities around. Whilst the measure of number of teacher strategies should relate to the pedagogy score then, we have low confidence in its utility in such a triangulation.

Teacher Confidence and Efficacy

Teacher confidence scores for 'generic' teaching skills increased by a statistically insignificant amount, as did confidence in the 6 areas of physics within the KS3 curriculum. Again, being able to discern which area of physics the respondents had engaged with would have allowed further investigation here.

Circumstantially, teachers reported a reduction in confidence after engaging with the literature as it highlighted the range of difficulties that pupils have in this physics. Conversely though, the above discussion of pedagogy suggests that teacher may also have gained a greater understanding of how to deal with these issues (although this is far from conclusive).

Reading Groups as a mode of Continual Professional Development (CPD)

Whilst the data is not reliable enough to draw conclusions from, the process of teachers and researchers meeting up to discuss pedagogical research was seen as incredibly positive by participants, as well as the researcher/tutors. One participant (from the chemistry design group) commented:

> *"I found the opportunity to read and discuss educational literature with researchers and other teachers a very valuable approach to CPD. It greatly enhanced my pedagogical understanding of chemistry teaching and increased my depth of knowledge of misconceptions in chemistry. Since completing the project, our Science Department have been working hard to increase the emphasis on teaching the 'language of chemistry' and to integrate a variety of strategies to promote active learning in our schemes of work for the new programme of study."*

Teachers commented that this form of CPD allowed them to focus on specific aspects of their subject teaching, and the research papers that each group read were selected upon the basis of developing interests. We feel that the ability of teachers to follow their own needs and interests, with the support of researchers, led to the qualitative success of the project (which is not clearly represented by the quantitative outcomes).

The design phase intervention worked best where teachers managed to protect time in order to engage with the reading and discussion. As will be discussed further in section 11, this was maximised when senior leaders 'bought into' the project and thus teachers were able to justify the process as part of their own development.

Another important realisation was that teachers were motivated by the concrete outcome of producing a scheme of work from the research, and we suspect that the impact of the process would have been considerably reduced if the teachers were simply reading the research without a clear relation to their planning.

From the researchers' perspective, teachers are able to quickly contextualise the research being considered. For example, whilst a research paper being read considered how density might be modelled using dots within a box to represent matter, the teachers quickly recognised that this would cause confusion later in relation to solids, liquids and gases. Teachers are constantly considering the holistic development of a child's understanding. There were other cases where research focused on very small groups of pupils, but teachers were able to see the limitations of this in relation to a whole class. Teachers often tried strategies from the literature before meeting in the reading groups, thus the discussions both informed but were also informed by practice. The experience of the project team throughout the project has allowed us to see that research informed pedagogies can only be developed by researchers and teachers working together in relation to the specific contexts in which research might be brought to bear on classroom learning. We intend to further develop this argument through future projects.

8.1b Teacher Outcomes – Trial Group

Date teacher intervention started: September 2014 (physics only)

Table 9b – Teacher Outcomes: teachers benefitting from the project

The 1st Return will either be your baseline data collected before the start of your project, or may be historical trend data for the intervention group. Please specify what the data relates to.

Target Outcome	Research method/ data collection	Sample characteristics	Metric used	1st Return and date of collection	2nd Return and date of collection
Increased subject knowledge: physics	Online survey: 6 questions taken from various research sources, designed to test conceptual understanding	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean score out of 12: two marks in each of the 6 areas of KS3 physics (percentage)*	*Group mean (percentage)=* **4.95** *(41%)* *Std. dev. =2.65* *n=21* *(Collected Sept-Dec 2014)*	*Group mean (percentage)=* **5.50** *(46%)* *Std. dev. =2.38* *n=3* *(Collected July 2015)*
Increased pedagogical skill: physics	Online survey: 6 questions taken from various research sources, designed to test pedagogic response. Follows Riese, Vigelsang & Reinhold (2011) classification of declarative and procedural Pedagogic Content Knowledge (PCK)	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean score out of 30: five marks in each of the 6 areas of KS3 physics (percentage)*	*Group mean (percentage)=* **4.19** *(14%)* *Std. dev. =1.78* *n=21* *(Collected Sept-Dec 2014)*	*Group mean (percentage)=* **11.00** *(37%)* *Std. dev. =3.90* *n=3* *(Collected July 2015))*
Increased confidence and efficacy: general teaching	Online survey: using standard survey developed by GLA	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean score based on a 1-9 scale of 'How much can you..'(1 – nothing, 3 – very little, 5 some influence, 7 – Quite a bit, 9 – A great deal)*	*Group mean =* **6.99** *Std. dev. =1.34* *n=21* *(Collected Sept-Dec 2014)*	*Group mean =* **6.18** *Std. dev. =1.67* *n=3* *(Collected July 2015))*
Increased confidence and efficacy: physics	Online survey: adapted from survey developed by GLA. 6 questions on each area of KS3 physics, one on practical work and one on uncovering misconceptions.	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean score based on a 1-9 scale of 'How much can you..'(1 – nothing, 3 – very little, 5 some influence, 7 – Quite a bit, 9 – A great deal)*	*Group mean =* **6.75** *Std. dev. =1.23* *n=21* *(Collected Sept-Dec 2014)*	*Group mean =* **5.91** *Std. dev. =1.61* *n=3* *(Collected July 2015))*

Number of strategies	Online survey question, asking for strategies that could promote conceptual change	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean number of strategies listed in open question*	*Group mean =* ***1.86*** *Std. dev. =1.35* *n=21* *(Collected Sept-Dec 2014)*	*Group mean =* ***2.50*** *Std. dev. =2.08* *n=3* *(Collected July 2015))*

Table 10b – Comparison data outcomes for Teachers

Target Outcome	Research method/ data collection	Sample characteristics	Metric used	1st Return and date of collection	2nd Return and date of collection
Increased subject knowledge: physics	Online survey: 6 questions taken from various research sources, designed to test conceptual understanding	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean score out of 12: two marks in each of the 6 areas of KS3 physics (percentage)*	*Group mean (percentage)=* ***4.95*** *(41%)* *Std. dev. =2.65* *n=21* *(Collected Sept-Dec 2014)*	*Group mean (percentage)=* ***8.00*** *(46%)* *Std. dev. =0* *n=3* *(Collected July 2015)*
Increased pedagogical skill: physics	Online survey: 6 questions taken from various research sources, designed to test pedagogic response. Follows Riese, Vigelsang & Reinhold (2011) classification of declarative and procedural Pedagogic Content Knowledge (PCK)	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean score out of 30: five marks in each of the 6 areas of KS3 physics (percentage)*	*Group mean (percentage)=* ***4.19*** *(14%)* *Std. dev. =1.78* *n=21* *(Collected Sept-Dec 2014)*	*Group mean (percentage)=* ***14.00*** *(67%)* *Std. dev. =2.00* *n=3* *(Collected July 2015))*
Increased confidence and efficacy: general teaching	Online survey: using standard survey developed by GLA	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean score based on a 1-9 scale of 'How much can you..'(1 – nothing, 3 – very little, 5 some influence, 7 – Quite a bit, 9 – A great deal)*	*Group mean =* ***6.99*** *Std. dev. =1.34* *n=21* *(Collected Sept-Dec 2014)*	*Group mean =* ***7.00*** *Std. dev. =1.27* *n=3* *(Collected July 2015))*
Increased confidence and efficacy: physics	Online survey: adapted from survey developed by GLA. 6 questions on each area of KS3 physics, one on	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean score based on a 1-9 scale of 'How much can you..'(1 – nothing, 3 – very little, 5 some influence, 7 –*	*Group mean =* ***6.75*** *Std. dev. =1.23* *n=21*	*Group mean =* ***7.61*** *Std. dev. =2.93* *n=3*

	practical work and one on uncovering misconceptions.		*Quite a bit, 9 – A great deal)*	*(Collected Sept-Dec 2014)*	*(Collected July 2015))*
Number of strategies	Online survey question, asking for strategies that could promote conceptual change	21 teachers responded to the pre-test and only 7 to the post- test (see 8.1.1b for discussion of sample)	*Mean number of strategies listed in open question*	*Group mean =* ***1.86*** *Std. dev. =1.35* *n=21* *(Collected Sept-Dec 2014)*	*Group mean =* ***2.00*** *Std. dev. =0* *n=3* *(Collected July 2015))*

8.1.1b Please provide information (for both the intervention group and comparison group where you have one) on:

- Sample size, sampling method, and whether the sample was representative or not
- Commentary on teacher impact (please also refer to table 5 re impact on different groups of teachers)
- Qualitative data to support quantitative evidence.
- Projects can also provide additional appendices where appropriate.

(Minimum 500 words)

<u>Samples</u>

21 teachers undertook an online pre-test and 7 the post-test. Of the post-test responses, the four in the trial group were from the same school, a comprehensive girl's school in central London. The three responses within the control group were made up of two responses from an independent school in Bromley, and one from a teacher in a comprehensive in Kent, who expressed an interest in the project (but was not funded). As with the design group, these tests were anonymised. Only one of the post-test responses can be matched to pre-test responses (for one of the control group teachers). There biographical information is summarised below:

Teachers Completing Physics Pre-test and Post-test in Trial/Control Group

	No. teachers	% NQTs	% Teaching 2 – 3 yrs	% Teaching 4 yrs +	% Male Female (withheld)	% Physics degree
Pre-test (combined)	21	14%	33%	54%	38% 52% 10%	10%
Post-test Overall	7	29%	29%	43%	50% 50% (0%)	14%
Post-test Trial	4	25%	25%	50%	71% 29% (0%)	25%
Post-test Control	3	33%	33%	33%	100% 0% (0%)	33%

We can thus only consider mean values for pre- and post-test from these groups, and must recognise that the post-test samples of trial and control groups are both so small as to be insignificant to analysis. This is furthered by the diversity of the settings in which they teach.

Commentary on Teacher Impact

As commented above, no conclusions can be drawn from the teacher data for the trial and control, due to the sample size, settings involved and the unavailability of matched data.

Comparison between Design Group and Trial/Control Group Data

One point worth speculating about is the increase in teacher physics knowledge and pedagogic skill scores between the trial and control group pre- and post-tests. These increases echo the increase in the values for the Physics Knowledge and Pedagogic Skill within the Design Group post-tests. Whilst no firm conclusions can be drawn, this would support the suggestion made earlier (section 8.1.1a) that the increase could be due to disparity in the tests, despite efforts to standardise them.

8.2 Pupil Outcomes

Date pupil intervention started: September 2014

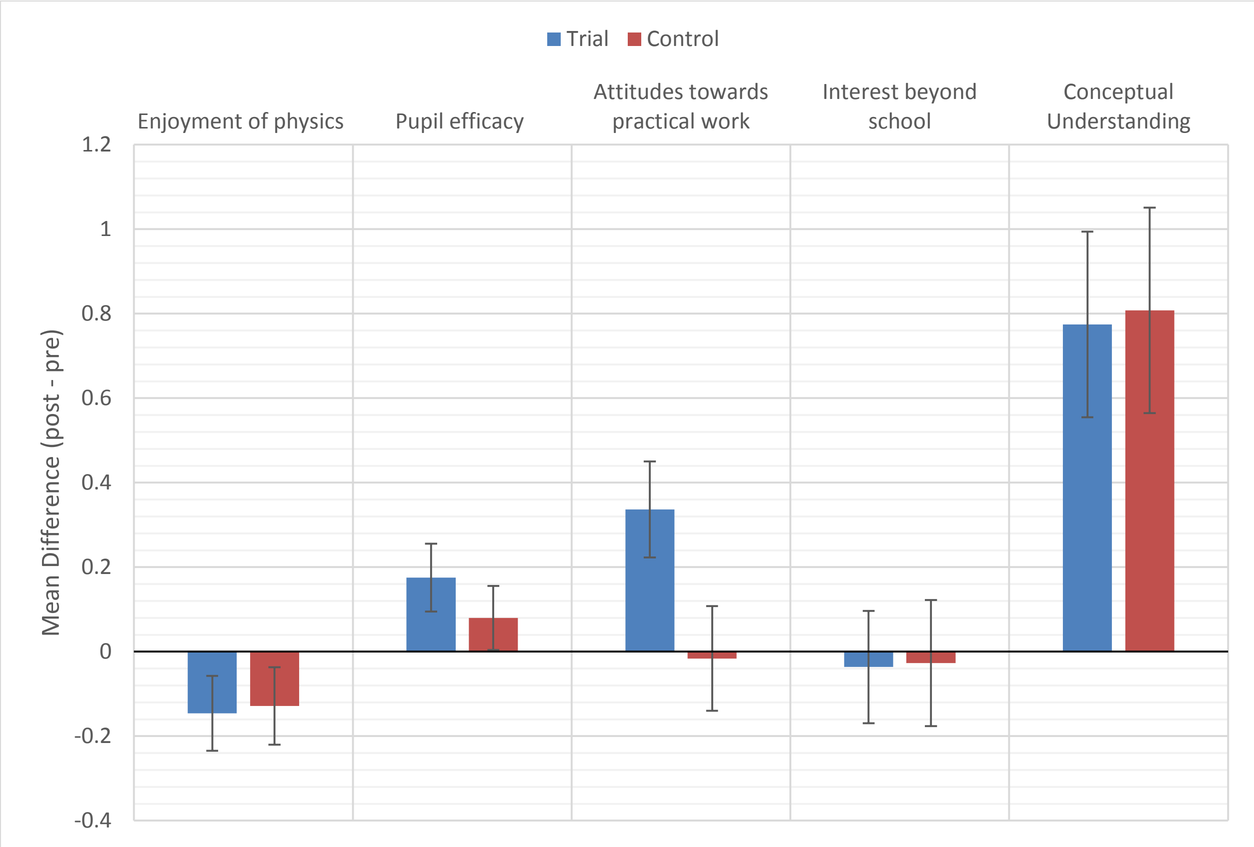
Table 11 – Pupil Outcomes for pupils benefitting from the project

Target Outcome	Research method/ data collection	Sample characteristics	Metric used	1st Return and date of collection	2nd Return and date of collection
Improved attitudes to Physics lessons	*Pupil self-report survey - Kind, Jones and Barnaby (2007) Science Attitudes survey (SAS)*	*Three schools returned a total of 214 pupils in the pre-test, of which 190 pupils could be matched in the post test.*	*Mean score on items for subscale 1 of SAS (enjoyment of physics at school) Possible score 1-5. High score=more favourable attitude*	*Group mean (std dev)= **2.8698** (.92865) n=214* *(Collected Jan 2015)*	*Group mean (std dev)= **2.7240** (.84863) n= 190* *(Collected June 2015*
Improved attitudes to Physics lessons	*Pupil self-report survey - Kind, Jones and Barnaby (2007) Science Attitudes survey (SAS)*	*Three schools returned a total of 214 pupils in the pre-test, of which 190 pupils could be matched in the post test.*	*Mean score on items for subscale 2 of SAS (pupil efficacy in physics) Possible score 1-5. High score=more favourable attitude*	*Group mean (std dev)= **2.5766** (.91792) n=213* *(Collected Jan 2015)*	*Group mean (std dev)= **2.7520** (.68495) n=190* *(Collected June 2015)*
Improved attitudes to Physics lessons	*Pupil self-report survey - Kind, Jones and Barnaby (2007) Science Attitudes survey (SAS)*	*Three schools returned a total of 214 pupils in the pre-test, of which 190 pupils could be matched in the post test.*	*Score on question 3 of SAS (attitudes towards practical work). Possible score 1-5. High score=more favourable attitude*	*Group mean (std dev)= **3.2311** (1.27624) n=212* *(Collected Jan 2015)*	*Group mean (std dev)= **3.5677** (1.0002) n=192* *(Collected June 2015)*
Improved attitudes to Physics lessons	*Pupil self-report survey - Kind, Jones and Barnaby (2007) Science Attitudes survey (SAS)*	*Three schools returned a total of 214 pupils in the pre-test, of which 190 pupils could be matched in the post test.*	*Score on question 4 of SAS (interest in physics beyond school: TV programmes/reading) Possible score 1-5. High score=more favourable attitude*	*Group mean (std dev)= **2.5573** (1.24776) n=192* *(Collected Jan 2015)*	*Group mean (std dev)= **2.5208** (1.35352) n= 192* *(Collected June 2015)*
Increased conceptual understanding in Physics	*Pupil responses to selected TIMMS questions*	*Three schools returned a total of 214 pupils in the pre-test, of which 190 pupils could be matched in the post test.*	*Total score across 12 selected TIMSS items. Higher score =better conceptual understanding*	*Group mean (std dev)= **5.5307** (2.08077) n=212* *(Collected Jan 2015)*	*Group mean (std dev)= **6.3051** (2.33485) n=195* *(Collected June 2015)*

Table 12 - Pupil Outcomes for pupil comparison groups

Target Outcome	Research method/ data collection	Sample characteristics	*Metric used*	1st Return and date of collection	2nd Return and date of collection
Improved attitudes to Physics lessons	*Pupil self-report survey -Kind, Jones and Barnaby (2007) Science Attitudes survey (SAS)*	*Three schools returned a total of 171 pupils in the post-test, of which 159 pupils could be matched to post test data.*	*Mean score on items for subscale 1 of SAS (enjoyment of physics at school) Possible score 1-5. High score=more favourable attitude*	*Control group mean (std dev)= **3.1685** (.74504) n=159 (Collected Oct 2014)*	*Control group mean (std dev)= **3.0397** (.91452) n=171 (Collected June 2015*
Improved attitudes to Physics lessons	*Pupil self-report survey -Kind, Jones and Barnaby (2007) Science Attitudes survey (SAS)*	*Three schools returned a total of 171 pupils in the post-test, of which 159 pupils could be matched to post test data.*	*Mean score on items for subscale 2 of SAS (pupil efficacy in physics) Possible score 1-5. High score=more favourable attitude*	*Control group mean (std dev)= **3.0401** (.62209) n=159 (Collected Oct 2014)*	*Control group mean (std dev)= **3.1196** (.75114) n=171 (Collected June 2015)*
Improved attitudes to Physics lessons	*Pupil self-report survey -Kind, Jones and Barnaby (2007) Science Attitudes survey (SAS)*	*Three schools returned a total of 171 pupils in the post-test, of which 160 pupils could be matched to post test data.*	*Score on question 3 of SAS (attitudes towards practical work). Possible score 1-5. High score=more favourable attitude*	*Control group mean (std dev)= **3.8875** (1.05799) n=160 (Collected Oct 2014)*	*Control group mean (std dev)= **3.8713** (1.19107) n=171 (Collected June 2015)*
Improved attitudes to Physics lessons	*Pupil self-report survey -Kind, Jones and Barnaby (2007) Science Attitudes survey (SAS)*	*Three schools returned a total of 171 pupils in the post-test, of which 160 pupils could be matched to post test data.*	*Score on question 4 of SAS (interest in physics beyond school: TV programmes/reading) Possible score 1-5. High score=more favourable attitude*	*Control group mean (std dev)= **2.7813** (1.27221) n=160 (Collected Oct 2014)*	*Control group mean (std dev)= **2.7544** (1.43831) n=171 (Collected June 2015)*
Increased conceptual understanding in Physics	*Pupil responses to selected TIMSS questions*	*Three schools returned a total of 169 pupils in the post-test, of which 162 pupils could be matched to post test data.*	*Total score across 12 selected TIMSS items. Higher score =better conceptual understanding*	*Control group mean (std dev)= **5.4938** (2.30041) n=162 (Collected Oct 2014)*	*Control group mean (std dev)= **6.3018** (2.11374) n=169 (Collected June 2015)*

Represented Graphically:



8.2.1 Please provide information (for both the intervention group and comparison group where you have one) on:

- Sample size, sampling method, and whether the sample was representative or not Commentary on pupil impact (please also refer to table 6-8 re impact on different groups of pupils)
- Qualitative data to support quantitative evidence.
- Projects can also provide additional appendices where appropriate.

(minimum 500 words)

See Appendix 1 for the analysis of pupil data.

<u>Samples</u>

The project engaged 9 trial schools and 10 control schools, who were self-selecting from the networks of contacts of the project team. Of these 5 trial schools and 4 control schools returned pre-test data. At the end of the trial, only three trial schools and 3 control schools returned matched data. This meant a trial sample of 212 pupils, and a control of 171 pupils from which to draw the analysis. Some data points were missing within this sample, resulting in the above analysis.

The samples are not 'randomised' due to schools self-selecting whether they wanted to be a trial or control school. This was the best way to recruit schools at the time, as some already had schemes of work in place (see discussion in section 11). It does however mean that the trial and control samples are not directly matched, and our analysis relies upon assessing the mean increase (and standard deviation) between pre- and post-tests.

It should be noted that there was a 'school effect' seen within one of the trial schools and one of the control schools, which saw lower scores in both attitudes and conceptual understanding across the tests. These could have been to do with the classes chosen to be part of the sample within these schools (for example lower attaining classes than in the other schools).

Impact on Pupil Attitudes

Trial schools were statistically lower (less favourable) on attitudes to physics than the control group initially (due to cohort differences) and then the same was still true at post intervention stage. The scores for enjoyment of physics and interest in physics beyond school both declined in both groups, with no statistically significant difference between the trial and control schools. This 'dip' in attainment and enjoyment is well reported within the literature (e.g. see Braund (2009)).

However, with respect to the efficacy in physics (how capable pupils feel they are at physics) the trial group changed their attitudes more favourably than the control group. Whilst this is only a small effect, and no simple conclusions can be drawn it supports our suspicion that the scheme of work produced is more pupil-led, revolving around pupils expressing their understandings more frequently.

A larger effect can be seen within the score pertaining to a question assessing the utility of practical work in learning physics. This declined slightly in the control group, but there was a modest increase in the trial group. Again, no strong conclusions can be drawn, but the scheme did focus upon pupils using simple activities to explore their own thinking.

Impact on Pupil Attainment

Both trial and control groups increased there conceptual understanding over the year for which the trial ran (2014-15). There is no statistically significant difference between these increases.

The pre- and post-test utilised equations from the TIMSS database, an internationally deployed set of standardised tests that are used to compare understanding in mathematics and science globally. Matched questions were used which asked very similar questions in the pre- and post- test; we asked two questions for each of the 6 areas of KS3 Physics, and a questions about practical investigations. This therefore gave us a high degree of confidence in our capacity to compare the scores on these tests. However, this was at the expense of more 'open' questions which would have allowed us to probe conceptual understanding further. It should be noted that assessments of conceptual understanding in the research literature involve much more rigorous tests, which would have not been possible through the 'testing at a distance' approach necessary for a large scale trial. It is therefore possible that pupils in the trial group have a deeper understanding than the control group, and this is simply not detected by these closed questions (of course, the converse is also possible). It can be said that the trial scheme does not appear to be worse than other schemes of work in allowing pupils to progress their attainment in physics.

Teacher Adaptation of Schemes

It is noteworthy that qualitative feedback from the trial schools was that the teacher modified the schemes of work in order to make them more appropriate to the specific classes they taught: in terms of engagement, challenge, literacy level, cultural differences etc. This is normal, good practice in schools and we supported it as a project team when contacted by teachers. However, this does bring into question the capacity of any statistical trial to assess a scheme of work. Within the literature, distinctions are made between intended, planned, enacted, assessed, and learned curricula (Porter, 2004; Kurz, et al., 2010), and this brings into question whether an intended curricula can be assessed when it is enacted in very different ways.

8.3 Wider System Outcomes

Table 13 – Wider System Outcomes

Target Outcome	Research method/ data collection	Sample characteristics	*Metric*	1st Return and date of collection	2nd Return and date of collection
Use of new resources by teachers/ schools outside the intervention group	*Website Downloads of Scheme of Work.*	*Google Analytics analysis of website data.*	*Page visits and location.*	*Website live from 1st June 2014.*	*4, 388 page views (3,631 site visits) up to 27th Sept 2015.*
			Number of downloads	*Downloads of scheme available 14th Sept 2015*	*78 downloads of the scheme of work up to 27th Sept 2015. (23 from USA, 8 from Japan, 6 from China, 4 from Germany, 24 from UK and 13 unidentifiable)*
	Contact from teachers/researchers	*Logging of e-mail communication.*		*26 schools in London and one in Manchester expressed initial interest when the trial was 'advertised', July 2014*	*3 further schools: 2 from London and 1 from Kent contacted us during the trial. We have had contact from researchers and teachers in Switzerland and Germany.*
Programme activities/ model is embedded in department/ schools planning beyond the intervention group (in other subjects within intervention schools and in other schools.)	*Feedback from design schools as to continuation of reading group format.*	*All design groups contacted.*	*Number of schools*	*None of the design schools had embedded at the start of the project (June 2013)*	*Two schools continue to meet for reading groups beyond the project funding. It should be noted that a further two of the chemistry groups are also still finalising plans through team meetings (Sept 2015).*

8.3.1 Please provide information on *(minimum 500 words):*

- Sample size, sampling method, and whether the sample was representative or not
- Commentary on wider system impact qualitative data to support quantitative evidence.
- Projects can also provide additional appendices where appropriate.

The intended impact of the project on the wider school system was twofold: firstly to support school based reading groups as an effective form of teachers' continuing professional development (CPD) and secondly to develop research-informed schemes of work which would be adopted by schools.

Uptake of Reading Group Model of Teacher Development

Of the 11 schools directly engaged in the design phase of the schemes, two are continuing with reading groups as modes of CPD currently (elicited from teacher contact September 2015). At the time of writing the report two of the chemistry design groups are still developing parts of the chemistry scheme through teaching it, and this may be seen as a further extension of the project process beyond the funded elements.

It is simply too soon after the project to see if there is a lasting influence upon the modes of CPD undertaken within the schools, and whether the physics and chemistry teachers involved will spread this model to other schools. However, we are currently planning two teacher engagement workshops in order to share the outcomes and learning from the project within London and the South East, which will allow us to promote this model further. We also plan to publish articles in academic and subject association journals in order to further this impact.

Interest in the Project and Downloading of Schemes of Work

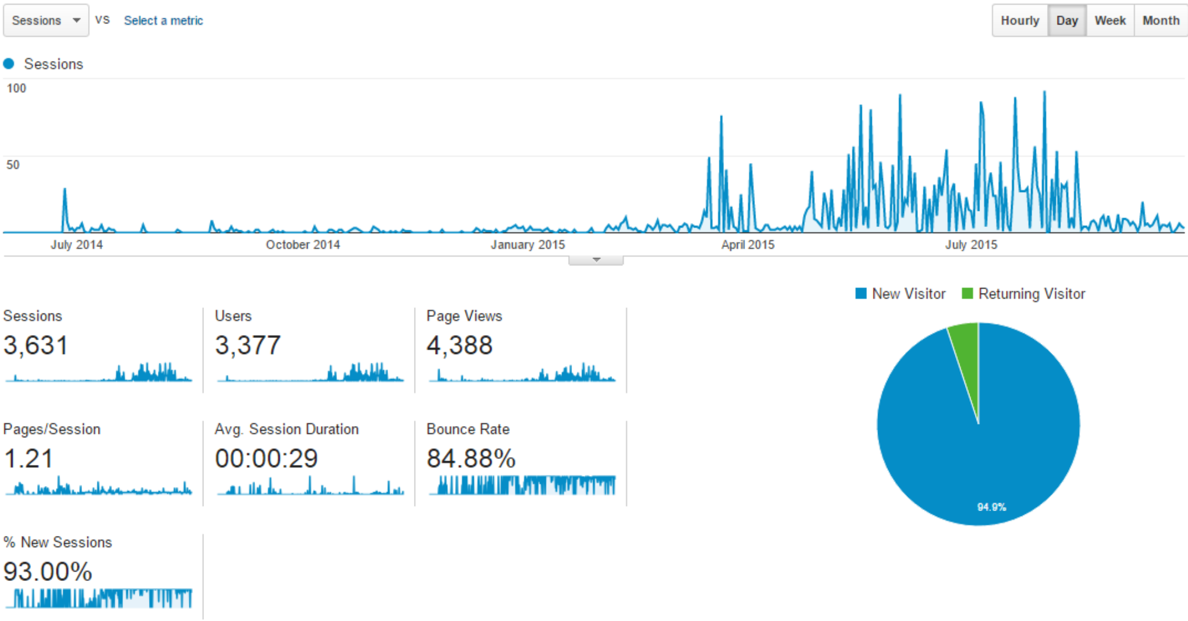
Because of the ongoing trial of the physics scheme of work, and our need to develop it following initial feedback, we did not post it to the project website for download until 14th September 2015. Just two weeks later there have been 78 downloads of the scheme of work, from around the world. Having only the country of origin data to go on, we are not able to ascertain what constitutes this interest however.

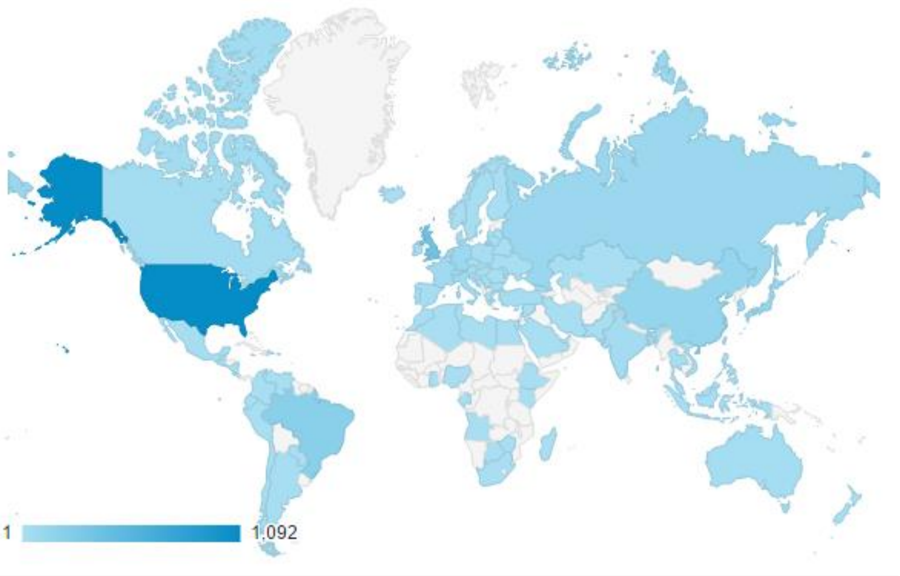
Through logging direct contact from schools we had contact from a total of 29 London schools, one in Kent and one in Manchester. We will share the outcomes of the project with them in line with publishing details of the teacher engagement events. Through presenting at the European Science Education Research Association conference in Helsinki, Finland in September 2015, we have had interest from two research groups: one in Germany and one in Switzerland, who are both interested in how conceptual change literature can be used with teachers. Our conversations with these groups continues and we shall explore future collaboration. We have also engaged with the Institute of Physics in the UK (who ran a separate LSEF project). We are working with them to explore further avenues of research and teacher engagement. Of particular interest to teachers and researchers alike has been a section of the project website which shares the 'misconceptions' which have been identified within the research literature. This list was compiled following the reading groups and continues to evolve, but the direct feedback from interested parties is that this is particularly useful.

Whilst again, we cannot discern the reason for interest in the project website, the team are pleasantly surprised that it has amassed 4,388 page views (3,631 site visits) up to 27th Sept 2015. This is especially positive when set in the context of the website containing only minimal information about the project, prior to the schemes being shared two weeks before writing this report. The below figure shows the variation of page views per day, since the site has been live. Whilst we can only speculate, the profile of visits around April, a pause for examinations in May, and then increasing traffic in the summer months is consistent with teachers using the site, and we hope is indicative of use in planning lessons.

Whilst circumstantial in nature, the profile of interest in the profile is very encouraging, especially considering that we have not undertaken any dissemination activity to date. We will continue to monitor interest both directly and online as we engage teachers both through the website (and social media) and through the established networks the project team inhabit.

Website views 1/6/14 – 27/9/15





8.4 Impact Timelines

Please provide information on impact timelines:

- At what point during/after teacher CPD activity did you expect to see impact on teachers? Did this happen as expected?
- At what point during/after teacher CPD activity did you expect to see impact on pupils? Did this happen as expected?
- At what point did you expect to see wider school outcomes? Did this happen as expected?
- Reflect on any continuing impact anticipated.

Teacher Impact Timelines

Whilst impact on the teachers within the design phase was expected by the end of that phase of the project, we anticipate continuing impact as those teachers go on to utilise the schemes of work that they have developed. This will allow teacher to apply their learning to their teaching directly. We invited teachers who had been in the design group to also engage in the trial, allowing a potential third group within the trial. After undertaking the pre- and post-test previously and having already tried out a number of the strategies and lessons with their groups, the measures deployed in the trial were no longer appropriate to this group. Nevertheless, we anticipate a greater continuing impact upon these teachers.

We anticipated the impact upon the teachers who taught the scheme of work within the trial to be diffuse, as we know that teachers use schemes of work in different ways depending upon their context and experience level. We anticipate this impact to continue as teachers embed their learning, and also utilise the schemes with different groups of pupils. However, we expect it to be diffuse such that it is immeasurable.

Pupil Impact Timelines

In line with discussion above, we anticipate a greater impact upon pupils taught by the design group teachers, than those taught by the trial group teachers, but this was not measured within the project design (this will be discussed further in the next section).

In anticipating that the impact on teachers will be diffuse, the impact upon pupils in control schools is also likely to be unpredictable. However, drawing upon a study in the USA (Gautreau & Novemsky, 1997) we were prepared for there to be a potential dip in performance on standardised tests when the trial group were first taught using a more ‘open’ conceptual change approach focusing upon exploring existing thinking. The study suggested that pupils did not cover as much material initially, but that once a deeper conceptual understanding is established they are able to accelerate beyond peers taught in a more traditional way. The results of the trial show no significant difference between the trial and control group in progression in standardised tests. Whilst the reliability of this result is unsubstantiated, it does suggest the scheme did not have a negative impact upon progression within a single year of teaching. Our engagement with the teachers in developing the scheme suggests that this may be because teachers did not wholly adopt an open, conceptual change approach in designing the scheme. We might speculate that teachers are bound by existing schools systems and their own confidence in existing approaches. Furthermore, teachers teaching the scheme in the trial group may not have adopted the approaches described, but once more modified them based upon their own school and experience.

9. Reflection on overall project impact *(maximum 1,500 words)*

In this section we would like you to reflect on:

- The overall impact of your project
- The extent to which your theory of change proved accurate
- How your project has contributed to the overall aims of LSEF
- Whether your findings support the hypothesis of the LSEF
- What your findings say about the meta-evaluation theme that is most relevant to you

Please illustrate using the key points from the previous detailed analysis.

All the evidence should be brought together here (achievement of outputs and outcomes, and the assessment of project impact) to produce well informed findings, which can be used to inform policy development in a specific area as well as the meta-evaluation of the LSEF.

This project demonstrated that the collaboration between teachers and educational researchers within reading groups is an effective mode of developing teachers. Qualitative evidence suggests that giving teachers access to research, time to engage with it and a specific output (in this case writing a scheme of work), allows them to respond to their own needs and develop their teaching. Educational researchers provide support in selecting, accessing and interpreting this research whilst teachers are uniquely situated to apply it to their own settings.

The absence of a control group means that measurement of the teachers engaged within this process should be treated with caution. Nevertheless, the teachers within this project advanced their physics knowledge score from a mean of 6.11 to a mean of 8.41. Their pedagogic skill score increased from 5.67 to 12.08. These are both statistically significant results which meet the aims of the project, although the reliability of the measures must be questioned. These quantitative results tentatively support the qualitative findings that teacher conceptual understanding of physics and chemistry and pedagogic understanding can be effectively developed through this approach to teacher research collaboration.

As well as developing the teachers involved, the project produced a research-informed scheme of work (a set of coherent lesson plans) for both physics and chemistry which has generated considerable interest within the educational community, even before it was released. 29 schools expressed a desire to be involved in the development and/or testing of the scheme from a single e-mail contact alone. Following presentation at a European conference, teachers and researchers from Germany and Switzerland have contacted the project team, and people from around the world have visited the website. The project website has received 4,388 hits up until 27th Sept 2015, and the scheme of work was downloaded 78 times within two weeks of it having been made available. This is all before any active dissemination activity (which is due to take place before January 2015).

This project was bold in seeking to test whether teachers who taught a research informed physics scheme of work, without being engaged in developing it, improved their knowledge and pedagogic skills beyond a control group of teaching using existing schemes of work. Furthermore, it sought to assess the impact upon pupils taught using the scheme. Because of difficulties in managing the trial 'at a distance', the project was only able to gather data from three schools who trialled the scheme (a total of 212 pupils) and three schools (a total of 195 pupils), who acted as a control by teaching existing schemes of work. This data suggested that enjoyment of physics and interest in physics beyond school declined in both groups, which is in line with existing evidence around pupils starting Key Stage 3. The conceptual understanding of both groups increased, with no significant difference between the two; this was measured by standardised, short answer questions rather than involved exploration of pupils thinking though, and should be seen as a narrow measure of attainment. Pupil efficacy increased slightly more in the trial group, but this was not statistically significant. The only significant difference was in attitudes towards practical work in physics, which appears to have increased in the trial group but declined slightly in the control group.

No simple conclusions should be drawn from these results because the sample size means that cohort differences were pronounced. Feedback from teachers suggested that most of them modified the schemes of work as they taught them, to make the appropriate to their students and teaching styles, but also to accommodate school requirements. We must also question whether the scheme produced was sufficiently different from existing schemes, such that the influence of conceptual change research is pronounced enough to change practice.

Overall, the trial shows that the research informed scheme of work which was developed within the project does not significantly affect pupil outcomes, with the possible exception of their attitudes towards practical work as important in learning physics. However, the impact on pupils over longer timescales cannot be discerned from this data, and we hypothesised that greater conceptual understanding may lead to accelerated learning over the coming years.

The London Schools Excellence Fund (LSEF) is based on the hypothesis that investing in teaching, subject knowledge and subject-specific teaching methods and pedagogy will lead to improved outcomes for pupils in terms of attainment, subject participation and aspiration.

The aims of the Fund:
I. Cultivate teaching excellence through investment in teaching and teachers so that attention is re-focused on knowledge-led teaching and curriculum.
II. Support self-sustaining school-to-school and peer-led activity, plus the creation of new resources and support for teachers, to raise achievement in priority subjects in primary and secondary schools (English, mathematics, biology, chemistry, computer science, physics, history, geography, languages).
III. Support the development of activity which has already been tested and has some evaluation (either internal or external), where further support is needed to develop the activity, take it to scale and undertake additional evaluation.
IV. In the longer term, create cultural change and raise expectations in the London school system, so that London is acknowledged as a centre of teaching excellence and its state schools are among the best in the world.

The need for longitudinal data means that the hypothesis of the LSEF cannot be fully tested within a project of this length. In terms of LSEF aims however, the project has supported the cultivation of excellent teachers through developing their knowledge and pedagogic skill together. In line with the established international research literature on Pedagogical Content Knowledge (PCK) (Shulman, 1986; Kirschner & Borowsi, 2011), we argue that knowledge-led teaching must include the development of pedagogic strategies. This project offers a mode of doing so through teacher-researcher collaborations.

A new scheme of work has been developed for physics and chemistry which embeds research evidence. Whilst the longitudinal impact of such a scheme is yet to be established it does not appear to limit pupil progression in the short term, and in the long term it may have a role to play in developing teaching and learning in these shortage subjects.

Furthermore, the linking of research and teaching in this way has the potential to influence the broader London system through peer interactions.

The project has provided tantalising suggestions as to the possibility of developing innovative and research-informed pedagogies, appropriate to the specific contexts of London schools. The project team intend to build on this by developing further activity which can be evaluated over longer timescales.

10. *Value for Money*

A value for money assessment considers whether the project has brought about benefits at a reasonable cost. Section 5 brings together the information on cost of delivery which will be used in this section.

10.1 Apportionment of the costs across the activity

Please provide an estimate of the percentage of project activity and budget that was allocated to each of the broad activity areas below. Please include the time and costs associated with planning and evaluating those activity areas in your estimates.

Broad type of activity	**Estimated % project activity**	**£ Estimated cost, including in kind**
Producing/Disseminating Materials/Resources	20%	£29,900
Teacher CPD (face to face/online etc)	40%	£72,500
Events/Networks for Teachers	10%	£2,566
Teacher 1:1 support	0%	£0
Events/Networks for Pupils	0%	£0
Trial of scheme of work	20%	£31,775
TOTAL	100%	£136,741 (same as total cost in section 5)

Please provide some commentary reflecting on the balance of activity and costs incurred: Would more or less of some aspects have been better?

The balance of costs is commensurate with the activity to complete each part of the project. Much of the budget went to partner schools to support teacher activity and this accounts for a large proportion of the costs in CPD, producing materials and the trial of the scheme of work. Management and direct staff costs of the project team are relatively small and are distributed across these areas. Events and teacher networks were important in establishing the programme and developing a consistent scheme of work, but their relatively low cost reflects the nature of these events in making strategic decisions as a group, rather than the much more involved work of evaluating research and writing schemes of work.

The trial arguably constitutes a larger proportion of the budget than is supported by its eventual utility as a source of evidence. However the trail had the potential to underpin much greater impact on the wider school system by providing quantitative evidence of the effectiveness of the approaches deployed.

10.2 Commentary of value for money

Please provide some commentary reflecting on the project's overall cost based on the extent to which aims/objectives and targets were met. If possible, draw on insight into similar programmes to comment on whether the programme delivers better or worse value for money than alternatives.

As discussed above, the trial might be considered as pertaining to research evidence in order to influence the broader educational system. If this is excluded then a total cost of approximately £100,000 was spent in the development of 40 teachers, equating to approximately £2,500 per teacher. Whilst this initially sounds like a large sum, it entailed the teachers evaluating several (at least 6) research papers, meeting with colleagues and researchers in school 12 times, attending two central workshops, interacting by e-mail with Professors from Imperial College and ultimately planning a series of lessons drawing upon all of this. It also includes the cost of covering that teacher's time and the facilities and resources that the involved institutions provided. Furthermore, this includes the purchasing of research papers and books, travel of the researchers, the development and maintenance of the project website, editing of the schemes of work and all the management and administration costs. The impact on the pupils that these teachers taught and their colleagues is also omitted from this analysis.

Undoubtedly, these costs could be reduced if there was greater efficiency in how the costs paid to schools were deployed in ensuring teachers were supported in undertaking the project, and this will be discussed in section 11. Overall, the project costs are reasonable for high quality teacher development which is tailored to those teachers and their settings, and which produces a research informed scheme of work and evaluation findings which have potential to impact the wider school system.

10.3 Value for money calculations

Note: This section is only required for projects with control or comparison groups

In order to demonstrate the cost effectiveness of the project we would like those projects who had control or comparison groups to provide some value for money calculations. Further guidance will be issued to support projects with this.

The main intervention within the project was the teacher reading and design groups, which did not have a control group. Therefore the relative value for money cannot be calculated.

11. Reflection on project delivery

This section is designed to allow for a discussion of wider issues relating to the project. (maximum 1,500 words)

Please include reflection on the following:

11.1 Key Enablers and Barriers to Achievement

- *Were there internal and/or external factors which appear to have had an effect on project success, and how were these responded to (if applicable)?*
- *What factors need to be in place in order to improve teacher subject knowledge?*

A key barrier to the success of the project was the ‘buy in’ of school leadership within the reading group and trial/control group schools. The project was established through using the existing networks of teachers across science departments which stem from the project team being engaged in teacher training in London. This meant that departments were recruited where one or more of the science teachers was known to the project team, or through word of mouth stemming from this. Whilst this allowed for good working partnerships, later in the project the teachers engaged in the reading groups, teaching the schemes or administering/engaging in the pre- and post-tests found that pressures on their time made the project difficult to continue. Despite funding towards teacher cover, the realities of school life, especially around the time of public examinations, meant that teachers were unwilling to leave their classes to others and/or they had competing pressures on their time, which were seen as more pressing by their colleagues and managers. The project succeeded only because the teachers involved worked within their own time to make it a success. The project team should, in hindsight, have worked to engage senior leaders further, or to have worked with a federation of schools which would ensure buy in at the senior level. Ideally, the project would have been established at the start of an academic year, making the reading groups a timetabled activity, accounted for in school budgets.

The development of teacher knowledge and pedagogic understanding/skill therefore developed most where teams of teachers had support from colleagues to engaging with and deploy research. This commonly resulted from school leaders recognising the value of teachers determining and addressing their own development needs and being given time to do so. Yet we found that teachers developed their understanding when this engagement with research was focused upon a specific output, in this case a scheme of work.

Furthermore, we found that teachers developed their subject knowledge best when they were able to follow their interests and needs within research literature, and this was supported by researchers who were able to select and interpret appropriate research papers in collaboration with the teachers.

11.2 Management and Delivery Processes

- *How effective were the management and delivery processes used?*
- *Were there any innovative delivery mechanisms and what was the effect of those?*
- *Did the management or delivery mechanisms change during the lifetime of the project and what were the before or after effects?*

This project was managed directly by the project lead, and this became difficult when other work pressures emerged. Administrative support was deployed ad hoc and in hindsight the project would have benefitted from the engagement of a dedicated project officer. In line with the above discussion around ‘buy in’ from school leadership, the engagement of schools for the trial and control groups was conducted primarily by e-mail, using already established, but diffuse, networks of contacts. The deployment of a dedicated officer to maintain relationships with these schools, as well as a greater level of engagement with leadership teams, would have improved return rates for data, and allowed the project to have greater influence.

As well as the management of the project, the delivery mechanisms involved the research team being able to find a mutually agreeable meeting time with design group schools. This would again have been furthered by meetings being an integral part of the school calendar.

A school or federation of schools wishing to deploy this effective form of staff development should consider engaging a project officer/coordinator to oversee it across schools, and engage researchers from local institutions to both support the groups, and carry out rigorous evaluation.

11.3 Future Sustainability and Forward Planning

- *Do you have any plans for the future sustainability of your projects?*
- *What factors or elements are essential for the sustainability of your project?*
- *How have you/will you share your project knowledge and resources?*

The project team are investigating avenues for further testing of the reading/design groups model of teacher development. Discussion have already taken place with the Institute of Physics (who ran a separate LSEF project) with a view towards together seeking funding for large scale testing of this approach.

The sustainability of reading/design groups within schools lies beyond the remit of the project team (who are university based). However, we are planning two teacher engagement workshops before January 2015, and a number of publications and conference presentations. This, along with the increasing interest in the project website are set to promote this model of teacher development, now that the research informed schemes are available for download and we have some initial evidence around impact.

12. Final Report Conclusion

Please provide key conclusions regarding your findings and any lessons learnt *(maximum 1,500 words)*.

Alongside overarching key conclusions, headings for this section should include:

Key findings for assessment of project impact

- *What outcomes does the evaluation suggest were achieved?*
- *What outcomes, if any, does the evaluation suggest were not achieved or partly achieved?*
- *What outcomes, if any, is there too little evidence to state whether they were achieved or not?*

A total of 11 schools and 40 teachers were engaged in reading groups and then the design of two schemes of work (two coherent sets of over 70 lessons): for KS3 physics and KS3 chemistry. The physics scheme was developed within 2013-14 and then tested in 2014-15, whilst the chemistry scheme was developed as a scaling up of the original project proposal (also in 2014-15).

Through collaboration with researchers, teachers were able to embed research evidence in these schemes of work, and within their own practice. Qualitative evidence and feedback from teachers supports the effectiveness of this approach to teacher development. This is further supported by the teacher scores on standardised tests (in physics only), which showed an increase in physics knowledge from 6.1 to 8.4, and in pedagogic understanding from 5.7 to 12.1. Whilst statistically significant, the absence of a control group reduces the reliability of this result. It does however correlate with the qualitative findings.

Teacher efficacy in general teaching, in physics in particular, and the number of strategies deployed to uncover misconceptions did not show any significant change. These results should again be treated with caution owing to the nature of the measures. Qualitatively however, teachers reported an increase in awareness of the issues pupils might have in physics, which both advanced their pedagogic skill but often reduced their confidence that pupils would benefit from traditional teaching methods. Taken together, this evidence supports the argument that teacher-researcher collaborations are an effective way to develop teacher knowledge and skill, whilst also developing research-informed resources.

The project also tested the impact of teaching a research informed scheme of work in schools which were not involved in its design. The problems of managing a trial ‘at a distance’ and without the ‘buy-in’ from school leaders meant that the participation rates in the trial were lower than hoped (despite strong initial interest). Samples of teacher data were too small to draw significant conclusions, although results for both trial group (three teachers) and control group (three teachers) echoes the findings for the teachers who designed the scheme.

Three schools who taught the physics scheme (a total of 212 pupils) and three schools (a total of 195 pupils) who acted as a control by teaching existing schemes of work. This data suggested that enjoyment of physics and interest in physics beyond school declined in both groups, which is in line with existing evidence around pupils starting Key Stage 3. The conceptual understanding of both groups increased, with no significant difference between the two; this was measured by standardised, short answer questions rather than involved exploration of pupils thinking though, and should be seen as a narrow measure of attainment. Pupil efficacy increased slightly more in the trial group, but this was not statistically significant. The only significant difference was in attitudes towards practical work in physics, which appears to have increased in the trial group but declined slightly in the control group.

No simple conclusions should be drawn from these results because the sample size means that cohort differences were pronounced. Feedback from teachers suggested that most of them modified the schemes of work as they taught them, to make the appropriate to their students and teaching styles, but also to accommodate school requirements. We must also question whether the scheme produced was sufficiently different from existing schemes, such that the influence of conceptual change research is pronounced enough to change practice.

Overall, the trial shows that the research informed scheme of work which was developed within the project does not significantly affect pupil outcomes, with the possible exception of their attitudes towards practical work as important in learning physics. However, the impact on pupils over longer timescales cannot be discerned from this data, and we hypothesised that greater conceptual understanding may lead to accelerated learning over the coming years. This is supported by existing research evidence.

Key lessons learnt for assessment of project delivery

- *What activities/approaches worked well?*
- *What activities/approaches worked less well?*
- *What difficulties were encountered in delivery and how could they be mitigated in the future?*
- *Were there any additional or unintended benefits (e.g. increases in student attendance as a result of an intervention aimed at teachers)?*

Collaboration with researchers allowed teachers to access and interpret research evidence which answered the needs that they identified within their own understanding and practice. Through being engaged in a specific outcome of relevance to their practice (developing a scheme of work), teachers were able to interpret and evaluate research evidence and apply it to their own practice. This approach is an effective form of teacher development. Furthermore, it is an effective form of developing research informed pedagogies which are appropriate to the specifics of school context. This is of great importance to the research community also.

The project was established by researchers who utilised their existing networks of contacts within science departments across London, and was managed by the project lead (as a small part of his broader role). This created significant barriers in maintaining project activity over the course of the school year and the success of the project is attributable to the tenacity and dedication of the teachers involved. The support of senior leadership teams would have embedded project activity within schools and thus allowed time and resource to be deployed much more effectively. A dedicated project officer would have been able to coordinate across schools and maintain relationships more effectively. Thus, engagement at the whole school level, rather than directly with existing contacts would have improved the outputs of the project and reliability of trial data.

Informing future delivery

- *What should the project have done more of?*
- *What should the project have done less of?*
- *What recommendations would you have for other projects regarding scaling up and/ or replicating your project?*

In hindsight, deploying a more qualitative and longitudinal approach to studying the effect of the reading groups would have yielded more useful analysis (but required a different level of funding and project timescale). The project sought to assess the impact of the scheme of work on schools not involved in its development, but could have paid more attention to the continuing impact on the groups of teachers who undertook the reading and design. The project analysis suggests impact upon their understanding and teaching ideas, but a longitudinal analysis, deploying a range of methods, would allow a more nuanced understanding of the reading groups on teacher practice and pupil learning. Any project undertaking to further the use of reading/design groups should consider this.

A range of other influences could be investigated in relation to developing teachers through their engagement with educational research. We are planning to study whether online reading groups and discussions achieve the same positive effects seen through face to face meetings, as well as investigate the role of networks in this interaction. The differences in developing subject knowledge and pedagogy in other subject areas could also be explored, as well as further testing the belief that permissive 'bottom-up' structures are conducive to more effective professional development.

As well as the need to engage school leadership and more closely maintain inter-school relationships in such projects, more sophisticated analysis of impact on pupil thinking could be deployed. We relied upon simplistic measures of attainment in order to use internationally benchmarked tests and to allow testing by teachers within a single lesson and 'at a distance'. More open, qualitative assessment is likely to expose the nuances of conceptual understanding, and allow a more detailed consideration of what influences it.

Bibliography

Please note that this is the bibliography used in writing the final report, considerably more papers were read by the reading groups within the project.

Anderson, C. & Smith, E. (1983) Children's conceptions of light and color: Understanding the concept of unseen rays. East Lansing: Michigan State University.

Babai, R. & Amsterdamer, A. (2008) 'The Persistence of solid and liquid naïve conceptions: A reaction time study, Journal of Science Education and Technology, 17 (6), pp. 553-559.

Braund, M. (2006) *Starting Science…Again? – Making Progress in Science Learning*. London: Sage.

Brown, D. E., & Hammer, D. (2008). Conceptual change in physics. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change*. New York: Routledge. (pp. 127-154)

Clement, J. J., & Steinberg, M. S. (2002). Step-Wise Evolution of Mental Models of Electric Circuits: A "Learning-Aloud" Case Study. *The Journal of the Learning Sciences*, 11(4), pp. 389-452.

DfE (2013) National Curriculum in England: science programme of study – key stage 3.

Duit, R. (2009), Bibliography: Students' and teachers' conceptions in science education. Kiel, Germany: IPN. Available at http://www.ipn.uni-kiel.de/aktuell/stcse/ [accessed 10th Nov 2013]

Gautreau, R., & Novemsky, L (1997) Concepts first: A small group approach to physics learning. *American journal of Physics*, 65, pp. 418-418

Heywood & Parker (2010) 'Cognitive Conflict and the Formation of Shadows' *The Pedagogy of Physical Science, Contemporary Trends and Issues in Science Education,* vol. 38, pp. 65-91

IOP: Institute of Physics (2010). Physics and: teacher numbers – An Institute of Physics Briefing note. Available at www.iop.org/news/10/sep10/file_44832.pdf [accessed 10th Nov 2013]

Kind, P. M. and Jones, K. and Barmby, P. (2007) 'Developing attitudes towards science measures', *International journal of science education*. 29 (7). pp. 871-893.

Kirschner, S.; Borowski, A., & Fischer, H. E. (2011) "Measuring Physics Teachers' Pedagogical Content Knowledge" in *Different Models and Methods to Measure Teacher PCK,* ESERA

Küçüközer, H., & Demirci, N. (2008). Pre-Service and In-Service Physics Teachers' Ideas about Simple Electric Circuits. *Eurasia Journal of Mathematics, Science and Technology Education*, 4(3), 303-311.

Kurz, A., Elliott, S. N., Wehby, J. H. & Smithson, J. L. (2010) Alignment of the Intended, Planned and Enacted Curriculum in General and Special Education and Its Relation to Student Achievement. *The Journal of Special Education*, 44(3), pp. 131-145.

National College of School Leadership, (2005). *Networked Research Lesson Study in practice,* Cranfield: CfBT.

Porter, A. C. (2004) Curriculum Assessment (Additional SCALE Research Publications and Products: Goals 1, 2, and 4), Nashville, Tennessee: Vanderbilt University.

Riese, J.; Vogelsang, C. & Reinhold, P. (2011) Pre-Service Physics Teachers' Pedagogical Content Knowledge in Different Teacher Education Programs. Proceedings of the ESERA conference, 2001. Availabe at www.esera.org/media/ebook/strand13/ebook-esera2011_RIESE-13.pdf [Accessed Sept 2015]

Scott, P., Asoko, H. and Driver, R. (1991) 'Teaching for conceptual change: A review of strategies', in R. Duit, F. Goldberg and H. Niedderer, Research in *Physics Learning: Theoretical issues and empirical studies.* Germany: University of Kiel, pp. 310-329.

Shulman, L. S. (1986) Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15:2, pp.4-14

Smith, E., Blakeslee, T. & Anderson, C. (1993) 'Teaching Strategies Associated with Conceptual Change Learning', *Journal of Research in Science Teaching*, 30 (2), pp. 111-26.

Vosniadou, S. and Brewer, W. (1990) 'A cross-cultural investigation of children's conceptions about the Earth, the Sun and the Moon: Greek and American data', In: H. Mandl, E. DeCorte, N. Bennett and H. Friedrich, *Learning and instruction: European research in an international context* (Vol. 2.2). Oxford: Pergamon, pp. 605-629.

Appendix 1 – Analysis of Pupil Data

Comparisons of control and trial group on attitudinal variables pre and post intervention

Group Statistics

	If trial or control school	N	Mean	Std. Deviation	Std. Error Mean
Pre test Mean Score Q1	control	159	3.1685	.74504	.05909
	trial	214	2.8698	.92865	.06348
Pre test Mean Score Q2	control	159	3.0401	.62209	.04934
	trial	213	2.5766	.91792	.06290
Pre test Score Q3	control	160	3.8875	1.05799	.08364
	trial	212	3.2311	1.27624	.08765
Pre test Score Q4	control	160	2.7813	1.27221	.10058
	trial	192	2.5573	1.24776	.09005
Post test mean score Q1	control	171	3.0397	.91452	.06994
	trial	190	2.7240	.84863	.06157
Post test mean score Q2	control	171	3.1196	.75114	.05744
	trial	190	2.7520	.68495	.04969
Post test score Q3	control	171	3.8713	1.19107	.09108
	trial	192	3.5677	1.10002	.07939
Post test score Q4	control	171	2.7544	1.43831	.10999
	trial	192	2.5208	1.35352	.09768

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Pre test Mean Score Q1	Equal variances assumed	15.886	.000	3.335	371	.001	.29866	.08955	.12258	.47475
	Equal variances not assumed			3.444	368.787	.001	.29866	.08672	.12813	.46920
Pre test Mean Score Q2	Equal variances assumed	33.474	.000	5.494	370	.000	.46349	.08437	.29759	.62940
	Equal variances not assumed			5.798	366.812	.000	.46349	.07994	.30630	.62068
Pre test Score Q3	Equal variances assumed	9.858	.002	5.279	370	.000	.65637	.12435	.41185	.90088
	Equal variances not assumed			5.418	366.713	.000	.65637	.12116	.41812	.89462
Pre test Score Q4	Equal variances assumed	.051	.822	1.662	350	.097	.22396	.13476	-.04108	.48900
	Equal variances not assumed			1.659	336.227	.098	.22396	.13500	-.04159	.48951
Post test mean score Q1	Equal variances assumed	.371	.543	3.402	359	.001	.31573	.09281	.13321	.49824
	Equal variances not assumed			3.389	347.740	.001	.31573	.09317	.13247	.49898
Post test mean score Q2	Equal variances assumed	3.039	.082	4.863	359	.000	.36758	.07558	.21893	.51622
	Equal variances not assumed			4.840	345.572	.000	.36758	.07595	.21819	.51697
Post test score Q3	Equal variances assumed	.624	.430	2.525	361	.012	.30364	.12027	.06712	.54015
	Equal variances not assumed			2.513	347.766	.012	.30364	.12082	.06600	.54128
Post test score Q4	Equal variances assumed	2.033	.155	1.593	361	.112	.23355	.14659	-.05472	.52182
	Equal variances not assumed			1.588	350.083	.113	.23355	.14710	-.05577	.52287

There was a significant difference between trial and control groups at pre and post intervention on q1, q2 and q3 of SAS. Controls were higher on all these scores.

Comparison of control and trial groups on attitudinal change between pre and post intervention for SAS subscales 1 and 2, TIMMS scores pre and post intervention and change in TIMMS scores

Group Statistics

	If trial or control school	N	Mean	Std. Deviation	Std. Error Mean
xq1meandifference (pre-post change in SAS q1)	control	121	-.2487	.87721	.07975
	trial	149	-.0238	1.30705	.10708
xq2meandifference (pre-post change in SAS q2)	control	121	.0006	.78503	.07137
	trial	150	.3200	1.08970	.08897
pre intervention score (TIMMS pre score)	control	162	5.4938	2.30041	.18074
	trial	212	5.5307	2.08077	.14291
Postestscore (TIMMS post score)	control	169	6.3018	2.11374	.16260
	trial	195	6.3051	2.33485	.16720
Testscorechange (pre-post change in TIMMS score)	control	146	.8767	2.35149	.19461
	trial	193	.7694	2.20321	.15859

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
xq1meandifference	Equal variances assumed	22.399	.000	-1.619	268	.107	-.22492	.13888	-.49836	.04852
	Equal variances not assumed			-1.685	259.317	.093	-.22492	.13351	-.48782	.03798
xq2meandifference	Equal variances assumed	16.711	.000	-2.707	269	.007	-.31945	.11801	-.55178	-.08711
	Equal variances not assumed			-2.801	265.793	.005	-.31945	.11406	-.54402	-.09487
pre intervention score	Equal variances assumed	2.003	.158	-.162	372	.871	-.03683	.22734	-.48387	.41020
	Equal variances not assumed			-.160	327.552	.873	-.03683	.23041	-.49010	.41644
posttestscore	Equal variances assumed	.707	.401	-.014	362	.989	-.00335	.23489	-.46527	.45856
	Equal variances not assumed			-.014	361.300	.989	-.00335	.23322	-.46200	.45530
testscorechange	Equal variances assumed	1.425	.233	.431	337	.667	.10728	.24879	-.38209	.59665
	Equal variances not assumed			.427	301.210	.669	.10728	.25105	-.38674	.60131

There was a significant difference between control and trial groups in the change in SAS q2 scores between pre and post intervention (trial group scored slightly higher) but no significant differences between the two groups in TIMMS scores pre intervention, post intervention or in terms of the change in TIMMS scores between pre and post intervention.

Attitudes to physics (across 3 out of the 4 measures used) were significantly different between the trial and control schools at both pre and post intervention but there was only a difference in terms of the change between pre and post intervention in one measure (q2) and impacts of the intervention on **knowledge** (as measured by TIMMS) were not significantly different between the trial and control groups at any stage.

Appendix 1

Evaluation Plan – LSEF159: Physics Foundations

Outputs	Indicators of Outputs	Baseline data collection	Impact data collection
a) Launch meeting during which project is set up and literature introduced.	Consensus achieved on division of scheme and process.	Attendees from each school (online sign up)	Register of attendance. Minutes of meeting. Analysis of churn.
b) Study meetings to engage with conceptual change literature (6 meetings). Supported by tutor & e-mail input from Physicists at Imperial College.	Notes from meetings show progression in conceptual understanding.	Initial conceptual understanding levels and confidence levels. Initial pedagogical understanding. All included in teacher pre-test, taken by all participating teachers.	Registers of attendance and notes from each meeting returned to tutors. Post-test taken by all teachers to assess conceptual understanding, confidence, pedagogical understanding (at end of development phase – July 2014).
c) Planning meeting in which format of scheme is decided	Consensus achieved on division of scheme and process. Standard format of plans & resources agreed upon.	Attendees from each school (online sign up)	Register of attendance. Minutes of meeting. Analysis of churn.
d) Meetings to plan 1/6 of the Scheme of Work. Supported by tutor & e-mail input from Physicists at Imperial College.	Teachers notes and lesson plans from meetings show application of conceptual understanding to planning. Scheme has clear links between conceptual change and activities.	Initial conceptual understanding levels and confidence levels. Initial pedagogical understanding. All included in teacher pre-test, taken by all participating teachers.	Registers of attendance and notes from each meeting returned to tutors. Post-test taken by all teachers to assess conceptual understanding, confidence, pedagogical understanding (at end of development phase – July 2014).
e) Scheme materials to be made available to trial schools with supporting resources and materials, so that it can be delivered by teachers.	Uptake of scheme by trial schools and collection of baseline data (Sept). Positive evaluation of scheme by physics teachers (outside development group).	Teacher pre-testing of conceptual understanding will be linked to teacher biographical and performance data (inc. years of experience and specialism). Pupil biographical data (including KS2 attainment) and school and class context data will be collected by trial schools. Pupil attainment data in conceptual knowledge test, will be linked to biographical data and teacher data.	Teacher post-testing of conceptual understanding will be linked to teacher biographical and performance data (inc. years of experience and specialism). Pupil biographical data (including KS2 attainment) and school and class context data will be collected by trial schools. Pupil attainment data in conceptual knowledge test, will be linked to biographical data and teacher data. Evaluation of scheme by teachers and pupils in trial.

Teacher Outcomes	Indicators of Outcomes	Baseline data collection	Impact data collection
Increased subject knowledge and greater awareness conceptual change approaches (all of development group)	Increased teacher scores in subject knowledge/ teaching approach tests Tests to be taken by all teachers involved in the development phase- We used questions taken from the literature on teacher Pedagogical Content Knowledge (PCK) supported by peer reviewed publications. Designed by project lead and reviewed by another Senior Lecturer in Education	Scores collected for individual teachers from pre intervention subject knowledge/ teaching method tests, linked to biographical data. Before start of project: April 2014	Scores collected for individual teachers from post development phase intervention subject knowledge/ teaching method tests, linked to biographical data. Qualitative data from session notes. July/August 2014
Increased teacher confidence in KS3 physics (development group)	Increased teacher scores in confidence surveys (teacher sense of self-efficacy) Survey to be completed by all teachers involved in the intervention	Scores collected for individual teachers from pre intervention confidence surveys Before start of project: April 2014	Scores collected for individual teachers from confidence surveys after development phase. Qualitative data from minutes of meetings (tutors attend fortnightly). July/August 2014
Increased subject knowledge, greater awareness of conceptual change approaches (trial group)	Increased teacher scores in subject knowledge/ teaching approach tests Tests to be taken by all teachers involved in the trial and control groups. This is the same test as given to the development group (see above).	Scores collected for individual teachers from pre teaching subject knowledge/ teaching method tests, linked to biographical data. Before start of trial: Sept 2014	Scores collected for individual teachers from post teaching subject knowledge/ teaching method tests, linked to biographical data. Qualitative data from teacher survey – sample of trial and control group. July 2015 (Analysis by Nov 2015)
Increased teacher confidence in KS3 physics (trial group)	Effect size on confidence relative to control group. Note: negative effect size may be linked to breaking down of misconceptions. Thus, linked to qualitative survey data. Schools in the control group are self-	Scores collected for individual teachers (intervention and control group) from pre teaching confidence surveys Before start of trial: Sept 2014	Scores collected for individual teachers from confidence surveys after teaching the scheme (intervention and control group). Qualitative data from teacher survey –all participants in trial and control group.

	selecting so teachers will be matched retrospectively across a range of dimensions: experience, subject specialism, school type (Fischer Family Trust data). We will also be matching pupils and teacher data so can match across pupil types (see below)		July 2015 (Analysis by Nov 2015)
Pupil Outcomes	**Indicators of Outcomes**	**Baseline data collection**	**Impact data collection**
Increased conceptual understanding of KS3 physics for all of trial group being taught with scheme.	Significant effect size in assessment above control group. All pupils in control group and trial group to sit test. Schools in the control group are self-selecting so pupils will be matched retrospectively across a range of dimensions (see next box). Pupils will also be matched to teachers of significance of teacher 'type' can be explored.	Pre-test of conceptual understanding of all pupils. Linked to data on KS2 attainment, LAC, FSM, FSM in 6 years, disadvantaged, EAL, gender, ethnicity, SEN. Before start of teaching physics in Sept 2014 Test questions are taken from TIMSS database 2011 & 2009 – questions used for international comparison of science understanding.	Post-test of conceptual understanding of all pupils. Linked to data on KS2 attainment, LAC, FSM, FSM in 6 years, disadvantaged, EAL, gender, ethnicity, SEN. Churn analysis. Data collected July 2015 Test questions are taken from TIMSS database 2011 & 2009 – questions used for international comparison of science understanding.
Pupil engagement with KS3 physics increased for trial group (on average).	Survey to evaluate scheme of work in trial and control groups.	Survey questions (incorporated in test) to assess engagement with physics Before start of teaching physics in Sept 2014 Survey is taken from Kind, Jones and Barmaby (2007)	Survey questions (incorporated in test) to assess engagement with physics. Sample of control and trial group to complete extended evaluation of scheme. Data collected July 2015 Sample: 20% of cohort in trial and control group: determined by biographical dimensions in pre-test to ensure representative spread of KS2 attainment, LAC, FSM, FSM in 6 years, disadvantaged, EAL, gender, ethnicity, SEN
School System / 'Culture Change' Outcomes	**Indicators of Outcomes**	**Baseline data collection**	**Impact data collection**

Use of new resources by teachers/ schools **outside the intervention group**	Uptake of new resources developed by LSEF programmes by non LSEF teachers/ schools	Planned new resources to be developed by LSEF programmes Avenues of dissemination/ promotion Dissemination for Sept 2014	Number of resources downloaded from websites (by different schools) Number of resources taken from training sessions/ conferences (by different schools) User feedback on quality of resources through online survey Data analysed after July 2015
Programme activities/ model is embedded in department/ schools planning beyond the intervention group (in other subjects within intervention schools and in other schools.)	Inclusion of programme activities/ model in development plans	Development plan pre roll-out of intervention Commitment/ sign up by school to specific criteria pre intervention	Part of department/ school/development plan Number of teachers following development plan/ due to roll out changes Commitment/sign up by school to specific criteria as part of project e.g. release of staff for 3 days of meetings and for study group meetings.