Evaluation of the impact of a continuing professional development (CPD) course for primary science specialists

Full Report

April 2015

wellcome trust

THE UNIVERSITY *of York*

Evaluation of the impact of a continuing professional development (CPD) course for primary science specialists

Final Report

Ian Abrahams

Judith Bennett

Alan Cheung

Louise Elliott

Pam Hanley

Zennifer Oberio

Peter Rudd

Robert Slavin

Maria Turkenburg

September 2014

CONTENTS

Executive summary

Context

This is the final report evaluating the impact of a continuing professional development (CPD) science course for primary school teachers. The course was developed by the Wellcome Trust and the National Science Learning Centre, and was designed to train primary teachers to become primary science specialists. Whilst the report mainly focuses on the impact of the course one year after its end, it also summarises findings from the start of the course and its end.

Methods

We ran a randomised controlled trial, using both quantitative and qualitative methods. The primary quantitative measures were teachers' and pupils' science subject knowledge; the secondary measures were teachers' confidence in the knowledge they were tested on, and pupils' attitudes towards science. The teacher measures covered the science specialists and also one nominated colleague in each school. We collected quantitative data at three stages: to provide a baseline at the start of the CPD course, at the end of the course, and to examine legacy effects one year after the end. These data were supplemented by a more qualitative evaluation, using case-study visits to a sub-sample of schools to assess implementation, effects on classroom practice and pupil reaction.

The baseline research initially involved 96 schools randomly assigned to three groups:

- **full CPD group** with science specialists receiving 14 days of directed CPD over a school year, with the equivalent of 10 further days of network support, online work and dedicated time in school
- **partial CPD group** with science specialists receiving four days of CPD over the school year
- **control group** with science specialists not receiving any of the CPD provided to either group above (but still able to access other CPD).

Attrition removed some schools from the study – not unexpected in a trial of such length and intensity. In all, 16 schools dropped out: eleven withdrew soon after random assignment and we removed five for not meeting the conditions of the study (either not teaching the correct Key Stage or not returning tests).

Of the remaining 80 schools, four were excluded from the legacy study looking at impacts in the year after the CPD had finished, because we did not receive pupil test data from the same individuals who had been tested at baseline. In total, the legacy comparisons with the baseline are based on test returns from 1582 pupils at 76 schools.

Data on pupils' attitudes were analysed for the 965 pupils who replied at all three stages.

The teacher sample was lower than the possible maximum of 76 specialists plus 76 nominated colleagues. This was because of a combination of factors, mainly staff turnover. In total, science subject knowledge tests were returned by 43 specialists and 35 colleagues, although because eligible tests had not been received from all the teachers at baseline, the analysis was based on 40 specialists and 32 colleagues.

We also conducted an extension study that involved sending pupil test papers and attitude surveys to those science specialists who took over a new Key Stage 2 class at the start of the academic year after the CPD course had finished. These schools were drawn from eligible volunteers from the main study and enabled us to follow any effects upon not only the previous classes but also the specialists' new classes.

Key findings

Analysis of the quantitative data showed no statistically significant impact of either the full or the partial CPD course on the teachers' test results or their confidence in answering the test questions. This was true of both the science specialists and their nominated colleagues. However, there are two caveats. First, as a result of under-recruitment of schools and attrition over the two years of the study, the sample of teachers remaining at the third stage may have been too small to show an effect at a statistically significant level. The findings, therefore, are indicative only. Second, the tests were previous Key Stage 3 science test papers (designed for students aged 13-14) and were deliberately not tailored specifically to reflect the topics covered by the CPD.

Despite this lack of quantifiable impact, almost all of the science specialists who gave their views – in case-study interviews or in self-evaluation forms completed at the end of the full CPD course – said that their subject knowledge had improved and that they felt much more confident about teaching science.

Multi-level modelling, carried out on pupils' subject knowledge scores across the three stages, found no statistically significant differences between the three groups. In other words, the subject knowledge scores of the full CPD group pupils were not significantly higher than those of the other groups. However, the extension study found that the scores of the full CPD group were significantly higher than those of the partial CPD group and those of both other groups combined. One possible explanation for this might be that the full CPD course ended in late June, with the order of topics covered being independent of the order in which teachers taught the topics to their classes. This may have meant that the teachers did not have the chance to revisit their course notes and materials and use these to improve their teaching until the following academic year.

Findings from primary science specialists

1. Participants' subject knowledge (and that of teaching colleagues also tested) was relatively weak, with an average baseline test result, on a Key Stage 3 paper, of only 59%.
2. Participants were generally positive about both of the CPD courses.

3. Participants in the full CPD course said that it had made them more enthusiastic about teaching science.
4. Participants in the full CPD group had different views about the value of the subject knowledge component.
5. The full CPD course had more reported impact than the partial course, though some participants in the latter felt that it provided the optimum balance between the benefits of the course and competing factors such as the practicalities of attending.
6. Participants in both CPD courses reported gaining new pedagogical subject knowledge and skills that improved both their performance as subject leaders and their classroom practice.
7. Whilst more than half of the full CPD group said that their subject knowledge had improved, this was not supported by the findings from the quantitative analyses.
8. Participation in the CPD was reported to have raised the status of, and increased support for, science in many of the full CPD schools and some of the partial CPD schools.
9. Participants in both CPD courses said that their teaching approaches had changed, to feature more practical, hands-on, open-ended, outdoor, and inquiry-based science activities.

Findings from pupils

10. The main study found no evidence that the full CPD course produced any statistically significant impact on pupils' science understanding. However, the extension study found that scores of pupils in the full CPD group were significantly higher than those of the partial CPD group and those of both other groups combined (although there were no statistically significant differences between the full group and the control group, or between the partial group and the control group).
11. There were a few instances where the surveyed opinions of pupils in the full CPD group appeared to have changed in a different way, or more extensively, than those of all other pupils.

Findings from schools

12. There was no evidence of CPD impact on colleagues' science subject knowledge.
13. A number of factors external to the CPD provision appeared to influence its impact, e.g. the support of the head teacher or the commitment of the teacher.
14. Views on the use of randomised controlled trials were comparatively neutral, and many participants did not have any idea what such a study design involved.

Overall, in terms of 'soft' measures, the CPD impacted positively on teachers of science in primary schools. The intervention raised teachers' confidence in teaching science and trying out new ideas, introduced teachers to new sources of teaching materials (such as the National STEM Centre) and also helped to foster and develop networks of primary science teachers who have remained in contact, sharing ideas and resources, beyond the lifetime of the intervention.

In terms of 'hard' measures the evaluation has found that the CPD (both the longer and shorter courses) had no statistically significant impact on teachers' subject knowledge or confidence in answering the tests. Results from the extension study found statistically significant differences in pupils' test scores, potentially indicating that some 'hard' measures of impact might require more time to manifest. Whilst some statistically significant changes were found in pupils' attitudes towards some very specific areas of science, the value of such changes would depend on whether they are transitory or enduring.

While feedback from teachers was generally positive, some considered the duration of the full CPD course to be too long. Such views, taken in conjunction with the 'hard' results of the RCT, lead us to suggest that for a national roll-out a shorter hybrid programme should be considered.

This study also demonstrates the importance of using a mixed-methods evaluation of an educational intervention in order to provide a mixture of both 'hard' and 'soft' evidence.

1. Background

The CPD course evaluated here was developed by the Wellcome Trust and the National Science Learning Centre. It was designed to train primary teachers to become Primary Science Specialists (PSS)[1]. For this evaluation, the course was aimed exclusively at those primary teachers who, whilst acting as science specialists within their schools, have no formal science qualifications beyond GCSE (or its equivalent) and also have no science-specific initial teacher training qualification. The principal aim of the course is to equip these teachers with the necessary subject-specific knowledge, pedagogical content knowledge (PCK), and leadership skills to successfully lead the teaching of science within their schools.

2. Objectives of the evaluation

The overall aims of the evaluation were to provide:

- a good quality evidence base of the school, teacher and pupil-level impacts of the CPD course after completion of the course;
- recommendations for the national roll-out of the initiative, if this is justified by the measured impact;
- a good quality evidence base of impacts one year after the completion of the course.

The evaluation was designed to capture the impacts on primary school pupils, science specialists and on schools generally, as specified below.

Impacts on primary school pupils:

- the science-specific attitudes and aspirations of pupils within their class and more widely within their school;
- the science achievements of pupils within their class and more widely within their school.

Impacts on science specialists:

- assess the science-specific subject knowledge of teachers and whether this improves as a result of the course;
- assess the scientific PCK of teachers and whether this improves as a result of the course;
- explore how the confidence of teachers is affected by the course;
- understand what factors, at the teacher level, can prevent full engagement with the course.

[1]The definition can be found in Appendix A.

Impacts of course attendance on schools:

- the status of science within the school;
- the school-wide development of science;
- the professional development of other teachers within the school who did not attend the course, including their PCK, science-specific subject knowledge and confidence.

3. Overview of the design of the evaluation

The research design had three principal sources of data.

- Pupil subject knowledge tests and attitude questionnaires that were undertaken three times over the duration of the study with the same pupils.
- Case-study visits that used semi-structured interviews with science specialists, other teachers of science and a member of the senior management team (generally the head teacher), a lesson observation and focus groups with pupils.
- Teacher subject knowledge and confidence assessments undertaken by science specialists and nominated other teachers of science three times over the duration of the study (i.e. pre- and post- CPD and the legacy year). These were always completed by the same teacher so if that individual was not available, the assessment could not be returned.

The data were collected in the context of a randomised controlled trial (RCT). This involved the random allocation of schools that had agreed to take part, and had signed a memorandum of understanding, into three equal-sized groups:

- **full CPD group** with school specialists experiencing 14 days of directed CPD[2] over a school year, and the equivalent of 10 further days comprising network support, online work and dedicated time in school;

- **partial CPD group** with school specialists receiving four days of CPD;

- **control group**, where school specialists did not experience any of the CPD provided.

Since the control schools were able to opt to take the CPD after the two years of the evaluation, this was a delayed treatment-control group design.

4. Scope of the report

The report compares initial baseline data with the outcomes collected throughout the evaluation as described above. In addition we report on an extension study that involved sending pupil test papers and attitude surveys to those science specialists who took over a *new* class of KS2 pupils at the start of the 2013 academic year. These schools were drawn from

[2] The CPD provided is described fully in Appendix B. This was not shared with the evaluators prior to the writing of this report.

eligible volunteers from the main study and enabled us to follow any effects upon not only the previous classes, but also the science specialists' new classes. The baseline tests and surveys for the extension schools were completed by October 2013 and the follow-up tests and surveys were completed in summer 2014 and, as such, there was only one post-test.

5. Methodology

5.1. Timing

The baseline study began at the end of the 2011/12 academic year. Teacher subject knowledge tests, including associated confidence measure indicators, were sent out to science specialists (also known as science subject leaders or co-ordinators) in all 96 schools that had agreed to participate and had signed a memorandum of understanding. The same tests were also sent to another teacher of science nominated by the head teacher (referred to in the report as the colleague). Equivalent tests of teacher subject knowledge with confidence measures were also administered at the end of the 2012/13 and 2013/14 academic years.

During September 2012, the baseline pupil tests and attitudinal questionnaires were sent out to science specialists in each of the three groups to administer to the class that they were teaching as part of this study. In the two subsequent stages, pupil papers were sent out alongside the teacher tests (i.e. summer 2013 and summer 2014) and were administered to the same set of pupils.

Whilst the vast majority of schools returned their baseline data relatively quickly there were a small number of schools where return of data was delayed, and it took over five months to receive full data sets. Hence the longer return periods for the baseline than for the two post-tests:

Stage	Timing	Papers
Baseline	July 2012-Jan 2013	Teacher papers
Baseline	Sept 2012-Jan 2013	Pupil papers
Post-test 1	June 2013-July 2013	Teacher and pupil papers
Post-test 2	June 2014-July 2014	Teacher and pupil papers

5.2. Quantitative measures

The teacher tests comprised both papers 1 and 2 of a past KS3 science test (re-titled A and B for the evaluation). Different KS3 tests were administered in 2012, 2013 and 2014 to prevent teachers predicting the question content, whilst maintaining a standardised level of difficulty. The maximum mark available was 150 on each occasion. The KS3 test was originally designed to be undertaken by students in Year 9 of secondary school (age 13-14). The questions are multi-part and mostly require write-in answers rather than ticking boxes.

Table 1. Distribution of marks on teacher test by science topic

Curricular area	Baseline	Post-test 1	Post-test 2
Sc1 (scientific inquiry)	33	42	39
Sc2 (life processes and living things)	41	35	38
Sc3 (materials and their properties)	38	37	37
Sc4 (physical processes)	38	36	36
Total	150	150	150

The proportion of marks available for each of the subject categories (as categorised by Qualifications and Curriculum Authority[3]) was stable across the years for Sc3 and Sc4, but there was more variation between Sc1 and Sc2 at baseline and post-test 1. A confidence rating was added to each full question, asking teachers to rate their confidence in their answers from 0 (total lack of confidence in all answers to this question) to 10 (extremely confident in all answers to this question).

In an attempt to make the analysis as robust as possible whilst optimising the sample size, only those teachers who had rated at least 14 of the questions (at both baseline and post-test 2) have been included in the following data. The cut-off point of 14 was chosen because this was around half the questions. Where teachers had failed to complete more than this, it was decided they could skew the results – for instance, they might only complete a confidence rating for the questions they found easiest.

Pupils completed tests designed for the project that spanned science curriculum content appropriate for the year group tested. Teachers were expected to administer these under normal testing conditions, with appropriate help given to any pupils with special support

[3] Qualifications and Curriculum Authority is no longer operational but archives can be accessed. www.qca.org.uk

needs. The pupil tests were constructed from a bank of standardised tests[4]. The papers were differentiated by year group, although five of the 20 questions on each test overlapped with the paper for the Year below and five with the paper for the Year above. The exception to this was Year 6. Because of a lack of suitable questions in the standardised tests, this paper was constructed using questions extracted from the 2009 KS2 science Standard Attainment Tests (SATs).

Post-test 1 papers were identical to those completed at the baseline, except for Year 3 pupils whose baseline results suggested there might be a ceiling effect after a further year (i.e. the level of difficulty would be insufficient to differentiate between the most able), so they were given the Year 4 test. For Post-test 2, pupils completed the test for the next academic year to retain age-appropriateness. Thus, a Year 4 pupil would complete a Year 4 paper at baseline and at post-test 1, and a Year 5 paper at post-test 2. The same was true for Year 5 pupils. However, because of the possible ceiling effect, Year 3 pupils would receive the Year 3 test at baseline; Year 4 at post-test 1 (labelled Year 3); and Year 4 at post-test 2.

Pupils also completed a simple, tick box attitude survey based on Pell and Jarvis (2001)[5]. The same version was used across all age groups and at each stage of the study. The survey consisted of 44 items on a Likert-type rating scale covering attitudes to school, to school science, to science as a subject, and to science in society. If children struggled with the language, teachers were invited to read statements out to them or explain the wording. For analysis, only those who responded to at least 40 of the 44 questionnaire statements were included in order to maintain a robust response set. (This is standard practice for this type of survey and reduces the amount of confusion that can result from missing data).

All the instrument measures used in the baseline and post-test 2 can be found in Appendix C.

5.2.1. Case Study visits

As part of the baseline study 30 schools (10 in each treatment group) were visited by five researchers, including a number of moderating co-observations. Recognising the pressure that schools are under we approached each school that returned a complete set of baseline data, asking if they would be willing to be a case study school, until each category was filled.

The purpose of these visits, following discussions with the Wellcome Trust, moved towards developing a broad overview of primary science education, including the effectiveness of primary science lessons and the role of the science specialist. In addition the visits have been used, at the request of the Trust, as an opportunity, given the potential growth in the use of

---

[4]We are grateful to Terry Russell and Linda McGuigan (Centre for Research in Primary Science and Technology, University of Liverpool) for permission to take these questions from Science Assessment Series 1/2.

[5]Pell, T., & Jarvis, T. (2001). Developing attitude to science scales for use with children of ages from five to eleven years. International Journal of Science Education, 33(8), 847–862. By kind permission of the authors.

RCTs in education, to gather information from teachers and head teachers on the use of RCTs in education in general as well as more specific issues relating to their own perceptions of being part of an RCT in one of the three treatment groups.

The case studies involved:

- observation of the science specialist teaching a science lesson
- interviews with the science specialist, the colleague, and the head teacher (or another member of the Senior Management Team (SMT) if the head teacher was unavailable)
- guided focus group discussions with two groups of five-to-six pupils, one from the science specialist's and one from the colleague's class.

Data collection techniques involved field notes and semi-structured interviews with members of staff as outlined in the previous paragraph. Generally speaking the compilation of the field notes used the same approach to assessing learning and the same evidence gathering processes that experienced science teachers use on a day-to-day basis as part of any Assessment for Learning (Black & Wiliam, 1998) exercise. Semi-structured interviews with teachers were used to probe in more detail the aspects explored in the on-line questionnaire, and, in particular, their views of the impact of the CPD on their classroom practice, on their pupils and more widely in their science department/school.

Case study visits were repeated between January and May 2014 to provide post-intervention comparisons.

5.3. Sample details

This section of the report focuses on a description of the sample for the RCT. It was important to establish how comparable the three different treatment groups were, especially since there was some attrition in the sample over the two years of the study.

Schools participating in the study were located in three geographical areas (London, Yorkshire and Humber, North West) as some CPD needed to be delivered at regional centres. The teachers participating as science specialists were all required to teach science in KS2 so for some teachers this meant teaching science to a different class during the school year 2012-2013.

5.3.1. Randomisation

The use of simple randomisation is problematic unless a study has a very large sample size, because the treatment groups may, by chance, be skewed on some characteristic that influences the findings. To reduce the risk of such chance bias, this study employed the matching technique described below.

In preparation for random assignment, all 96 participating schools were placed into matching triplets on the basis of attainment, percentage of pupils eligible for free school meals (FSM) and year group being taught science by the specialist (as far as was known at this stage of the study). Attainment was based on the average (mean) percentage of pupils achieving level 4 or above at KS2 in English and mathematics SATs over the last three years. FSM acts as a proxy indicator of a school's socio-economic context, and the year groups were Years 3, 4 and 5. One school from each triplet was then randomly assigned to each of the three groups: A, B, and C. Each treatment group therefore comprised 32 schools.

5.4. Attrition

Across the two years, as would be expected in a longitudinal study, there was some attrition of the sample (see Figure 1 on page 16). Soon after randomisation, 11 of the 96 schools withdrew, particularly from the full CPD and control groups. The reasons given for withdrawing from the study varied, as summarised below:

full CPD group (6):

- CPD is too much commitment (3)
- specialist off sick (1)
- personal circumstances (1)
- no details provided (1)

partial CPD group (1):

- no details provided (1)

control group (4):

- no longer willing to teach KS2 (1)
- SS left, other staff unwilling (1)
- no details provided (2)

We would emphasise that 'no details provided' has only been used here after repeated attempts to get a response from the school regarding the reason for their withdrawal went unanswered. Three more schools withdrew during the first year. Two were in the full CPD group: one was unable to commit to the CPD and the other specialist had performance issues which made the school unwilling to release her for the time necessary. In the partial CPD group, one school withdrew when both the head and science specialist left the school on its conversion into an academy. All three schools were retained in the analysis because they had already begun to engage with the CPD course and were therefore encouraged to provide data for the remainder of the evaluation.

Retaining schools even after they have withdrawn from the intervention is called "intention to treat" (ITT) analysis. This design is used to reduce the risk that the RCT might give artificially positive results if schools that have dropped out because they feel the intervention is ineffective are discarded from the data, potentially leaving an increasingly favourably-skewed active sample. Studies using an ITT design typically attempt to collect data from all schools that drop out at any stage after randomisation. In this study, it was decided on pragmatic grounds – because providing data would be very onerous for schools due to the large amount required across two years, and they had not even provided baseline pupil data at this stage – that only schools that had attended some training before withdrawing would be asked to remain in the evaluation as ITT.

As a result of the dropouts (11), and a further school being withdrawn from the study by the evaluators for non-compliance (the specialist was not teaching science to the KS2 class as expected), 84 schools were contacted for data at post-test 1.

At post-test 2, 80 schools were contacted, four fewer than post-test 1. Two of these four had failed to provide useable data at either of the previous stages (both in the full CPD group) and the other two were ITT schools who had failed to engage after the baseline despite numerous requests to do so. However, the third ITT school provided data at all three stages so it remained in the evaluation even though it had formally withdrawn from the CPD.

In total, across the two-year timeframe of the RCT, 16 schools withdrew, or were withdrawn, from the study. The initial rate of attrition was very high (11 out of 96) suggesting that participants had not fully understood or bought into the implications of random allocation and the requirement to remain in the study regardless of which treatment group they were placed into. There was some evidence that schools had not thought through the ramifications for the school of being assigned to the full CPD group (which demanded a lot of release time from the classroom) and subsequently withdrew because it was too much commitment. Others had not appreciated the key role of being in the control group with three dropping out (or having to be withdrawn) because they were not prepared to teach science to a KS2 class when their usual class was KS1.

In all, 19 science specialists in the remaining schools were teaching science to a non-regular class so as to be eligible for this study: 5 in the full CPD group, 6 in the partial CPD group and 8 in the control group. Others, wrongly, interpreted being placed into the control group as being the end of their involvement with the trial, and had to be encouraged to remain in the sample and return data (successfully in all cases). Some schools said they had received the impression at recruitment that the odds of them being in what the head teacher considered to be an ‘undesirable’ treatment group were relatively low (one in three) and therefore they should sign up anyway. These problems could potentially have been mitigated, to a greater extent than was the case, by giving the schools more information about RCTs through a recruitment conference. Indeed, whilst the evaluators were informed that an information leaflet was provided to all head teachers about RCTs, along with the memorandum of understanding that they were required to sign, it was clear, from the lack of awareness about the nature of RCTs and what they entailed, that the leaflets had either not be widely read and/or their information had not been assimilated. The large initial drop-out after schools were allocated to the three groups emphasised that it is critical, for all those who sign up to an RCT, to be willing to remain involved regardless of allocation.

Other factors may have contributed to attrition. These included having to randomise before pupil baseline data was collected due to constraints imposed by the time frame of the CPD course (this reduces the chance of keeping schools in the sample as ITT, see above). Each data collection stage involved a heavy burden of testing – teachers had to find an hour for their class to complete their papers; find an hour and a half to do their own test (which involved exposing their science knowledge – or lack of it - to scrutiny); and persuade a colleague to do the same. Some teachers had to complete the tests outside working hours, in other schools special arrangements had to be made to release the teacher for the requisite time. Despite prior warning in the memorandum of understanding, it is always difficult to find sufficient non-contact time in busy schools, especially with the culmination of activities at the end of the school year. It was not possible to incentivise the evaluation process by offering teachers useful feedback on their results or those of their pupils during the two year evaluation in case it contaminated the findings.

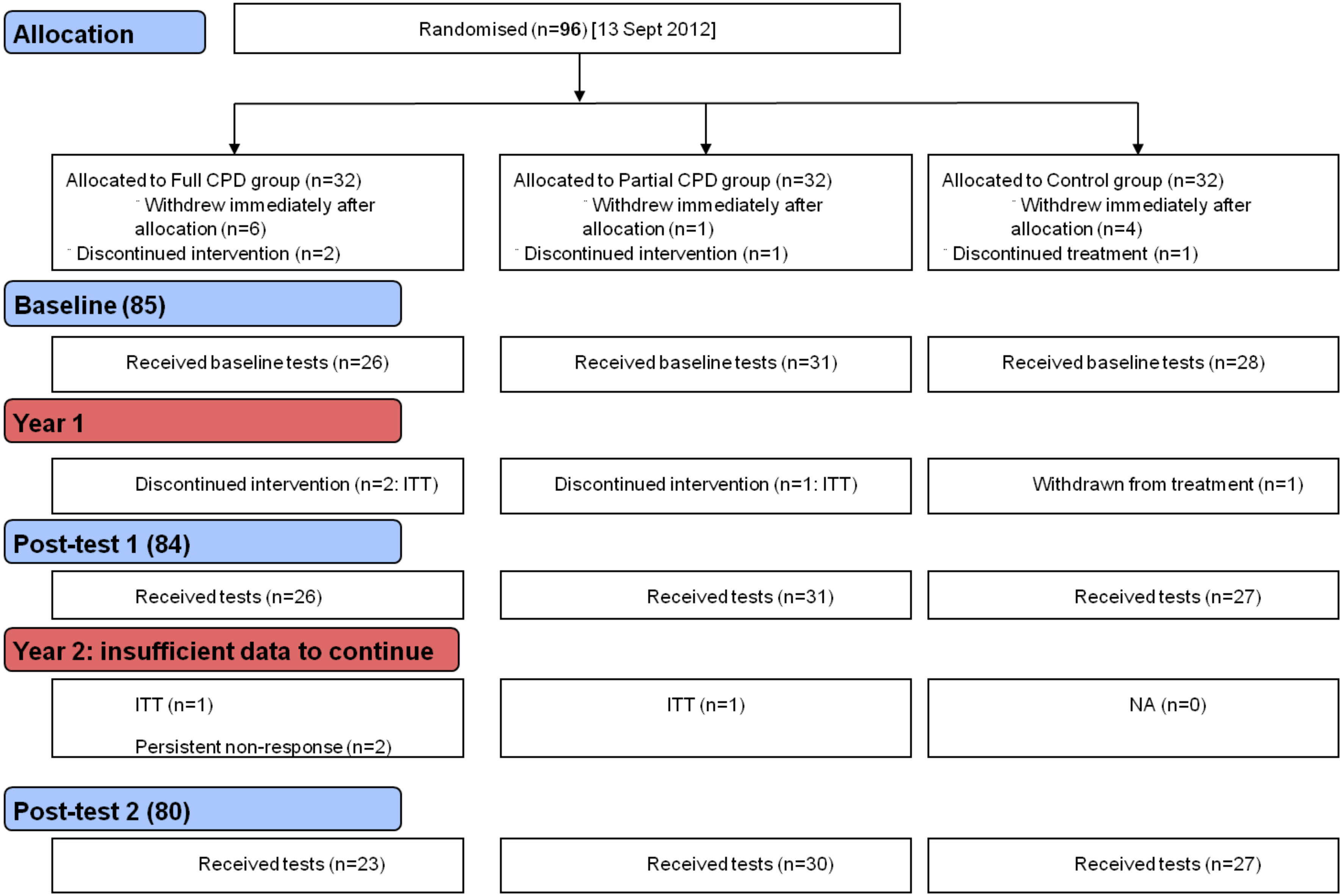
After the initial fallout, retention was good for a two-year study. The evaluation design followed best practice and, for example, used a wait-list i.e. all control schools were promised the treatment (or its equivalent) after the study was finished, albeit this was two years into the future. Furthermore, the science specialist received a financial incentive for returning data at each collection point, in the form of a £40 Amazon voucher. Also, a termly electronic newsletter was sent to all schools to develop and maintain a sense of connection with the evaluation team and to provide information about the evaluation, team members, contact details for the project administrator and reminders of upcoming data collection periods when relevant.

A further four schools failed to respond at post-test 2 leaving a final sample of 76 schools, equivalent to an overall attrition rate of 21%. To put this in perspective, the average attrition rate across projects completed under the auspices of the Education Endowment Foundation (these are mostly RCTs although some of much shorter duration), is 15%[6]. Since no bias is evident in the final sample compared with those originally recruited (see Section 7.1), we can be confident that the rate of attrition has not impacted on the findings.

---

[6]http://educationendowmentfoundation.org.uk/uploads/pdf/Classifying_the_security_of_EEF_findings_FINAL.pdf

Figure 1. Participation of schools throughout trial



6. Summary of the key findings from Baseline to Post-test 1

6.1. Baseline November 2012

6.1.1. Teacher test summary

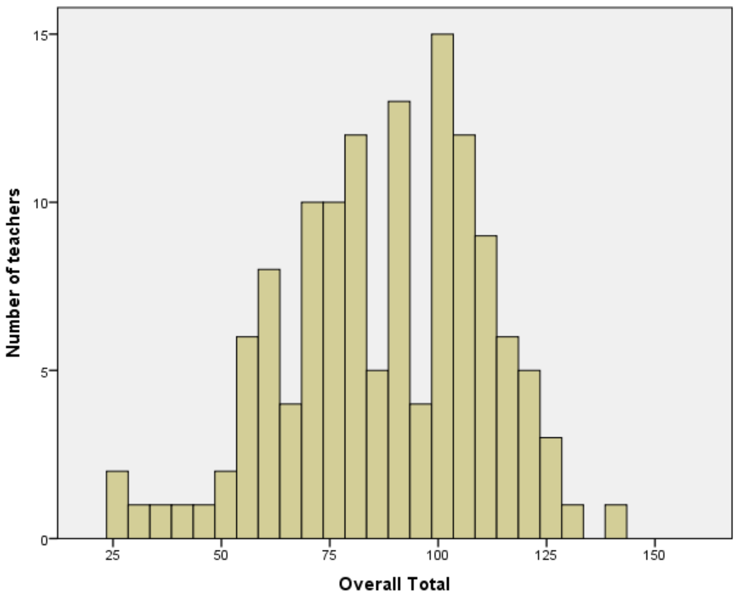
Teacher subject knowledge tests (a past KS3 science paper) and associated confidence measure indicators were sent out to science specialists, and another teacher of science nominated by the head teacher, at the end of the 2012/13 academic year. Tests were returned by 68 of the 85 schools remaining after randomisation: a full set of papers from 64 schools, specialist only from one and colleague only from three (see Table 2). Response rates for the partial CPD and control groups were around 80%. The full CPD group was lower at 67%. This apparent anomaly was primarily due to teacher tests being discounted if completed after starting the CPD programme as it might boost their subject knowledge and make their performance invalid as a baseline. Some test papers were also lost in the post or in the school.

Table 2. Number of teacher tests returned (science specialist and colleague)

	Full CPD group	Partial CPD group	Control group	Total
Teacher tests	35	52	45	132
Expected total	52	62	56	170
Response rate	67%	83%	80%	78%

The maximum mark available on the subject knowledge tests was 150. The average mark achieved was 88 (59%), although this covered a wide range - from a minimum of 26 (17%) to a maximum of 143 (95%). The graph below (Figure 2) shows the spread of scores.

Figure 2. The spread of scores from the teacher tests



A series of analyses comparing the test scores by different sub-groups of teachers was conducted at the baseline. There were no statistically significant differences between the performance of teachers in the three treatment groups. The average marks were 56% (full CPD group), 61% (partial CPD group) and 57% (control group). Nor was there any statistically significant difference in the marks achieved by the science specialists compared with their teacher colleagues (57% and 60% respectively).

All but three of the teachers were identifiable by gender. Male teachers (N=23) achieved an average mark of 57% and females (N=106) achieved 59%, which was not a statistically significant difference.

The questions were analysed by the four main categories of the primary science curriculum (Sc1 – scientific inquiry; Sc2 – life processes and living things; Sc3 – materials and their properties; Sc4 – physical processes) using the distribution of marks in the original Mark Scheme. Where different parts of a question were attributed to different categories (only a minority of cases), the question was allocated to the category accounting for the most marks. Teachers performed significantly better on Sc1 (66% average mark) than Sc2 (61%), and performance on both these categories was significantly better than for Sc3 (53%) and Sc4 (56%). This indicates that teacher subject knowledge was weaker in the area of physical science.

Teachers were asked to rate their confidence in their answers to each question they tackled, and only those teachers who had completed at least 14 confidence ratings across the 29 questions have been included in the following data (123 teachers out of the total of 132).

The mean confidence rating across both papers was 4.7, although the range was very large – the most confident teacher had a mean of 7.7, and the least confident only 0.1 (i.e. they had given 0, “a total lack of confidence”, to most questions they answered).

Teachers in the three different treatment groups showed no significant statistical difference in their confidence ratings, nor was there a significant difference between science specialist (4.5) and teacher colleague (4.9). Although not a statistically significant difference (due to unequal sample sizes of 22 men and 98 women), there was an indication that male teachers expressed more confidence than females (5.2 versus 4.6) despite there being no significant difference in achievement on the tests.

Confidence was analysed by science category (Sc). This shows that teachers were much more confident about Sc1 (scientific inquiry: 5.3) and Sc2 (life processes and living things: 5.3) than they were about Sc3 (materials and their properties: 4.0) and Sc4 (physical processes: 4.3). This reflects the differences in achievement reported above.

Overall, confidence ratings were positively correlated with performance ($r=0.54$, $N=123$, $p<0.01$). That is, the general pattern was for teachers to be more confident on the questions they answered correctly, and less confident on those they did poorly on.

This significant positive correlation between achievement and confidence ratings was evident for all four Sc categories. The correlation was strongest for Sc3 (0.59), and middling for Sc1 (0.47), Sc2 (0.5) and Sc4 (0.52).

6.1.2. Pupil test summary

The baseline pupil tests and attitudinal questionnaires were sent to all schools in September 2012. The baseline analyses were based on test papers from 1,980 pupils in 77 schools. The full breakdown of pupil test numbers, cross tabulated by year group, geographical region and treatment group is provided in Table 3 below. It can be seen that slightly more pupils are from Year 3 (739) than Year 4 (614) and Year 5 (627), but this skew towards Year 3 reflected the composition of the original sample.

Table 3. Numbers of pupils taking tests by region, year group and treatment grouping

Treatment			Year Group			Total
			Year 3	Year 4	Year 5	
Full CPD group	Region	Yorkshire & Humber	66	58	57	181
		London	87	83	45	215
		North-West	101	40	57	198
	Total		254	181	159	594 (21)[7]
Partial CPD group	Region	Yorkshire & Humber	20	65	55	140
		London	141	68	50	259
		North-West	104	68	146	318
	Total		265	201	251	717 (29)
Control group	Region	Yorkshire & Humber	51	77	96	224
		London	86	77	77	240
		North-West	83	78	44	205
	Total		220	232	217	669 (27)
Total	Region	Yorkshire & Humber	137	200	208	545
		London	314	228	172	714
		North-West	288	186	247	721
	Total		739	614	627	1980 (77)

[7] Number of schools included in brackets.

The average percentage marks for each year group, across all the treatment groups, and the ranges of marks are shown in Table 4.

Table 4. Pupils' average percentage marks for each year group and the range of marks

Year	Mean score (%)	Minimum (%)	Maximum (%)
Y3 (N=739)	65	7	96
Y4 (N=614)	47	7	84
Y5 (N=627)	39	2	85

The decrease in average score as the pupils increase in age suggests that Y4 and Y5 pupils found the difficulty level of their respective tests higher than the Y3 pupils. As this affects each treatment group equally, it will not be an issue for future analyses.

Since these were the first rounds of tests, establishing a base line for future tests of pupils in the three treatment groups, it was important to examine whether there were any significant differences in test outcomes across these three groups. In order to do this pupil tests scores were standardised, so that we could compare scores across the three year groups. Detailed analysis revealed that there were **no statistically significant differences** (p=<0.05) in the test results of the three treatment groups: a one-way between-groups analysis of variance showed no significant differences between treatment groups. Raw scores, standardised z-scores and standard deviations were as follows: full CPD group: 29.6, z=0.001 (SD=7.24); partial CPD group: 28.3, z=-0.077 (SD=7.12); control group: 29.67, z=0.817 (SD=7.45). F (2, 736) = 2.88, p = 0.06.

Analysis of the baseline results by pupil gender was also carried out. The tests were completed by 980 boys and 950 girls (with gender omitted or not stated on 50 papers). The mean scores achieved were 49% for boys and 53% for girls and this difference was statistically significant (at the level of p<0.01)*:* An independent samples t-test showed a significant difference in percentage scores for boys (M=49.01, SD=18.95) and girls (M=53.39, SD=18.63; t(1928)= -5.11, p=0.000, two-tailed).

6.1.3. Pupil attitudes

There were 44 questions divided into four sections (see below). Completed questionnaires were returned from 2,033 pupils. However, questionnaires with more than three missing responses were excluded, leaving a sample of 1,782 for analysis. The focus in this analysis was on change in pupil attitudes. The section on post-test 1 pupil attitudes (Section 6.2.3 below) outlines the most important changes that were identified between baseline and post-test 1.

6.1.4. Case study observations

One of the points to emerge from the case-study visits undertaken was the extent to which there is clear evidence for well-planned effective science practical work taking place across all three groups although we would caution that schools taking part in this study and consenting to case study visits in any group might have been ‘pro-science’. While science specialists might lack specific science qualifications beyond GCSE (or its equivalent) there was clear evidence of pupils both ‘doing’ and ‘learning’ with both objects and ideas (Abrahams & Millar, 2008).

One factor contributing to the effectiveness of these lessons, and mentioned by some science specialists in the sample, has been the extent to which they have received support in the form of Local Authority (LA) (Hackney, Tower Hamlets and Hounslow) school networks designed to provide CPD for science specialists. Similarly science specialists, in two out of three schools visited in Cheshire, also spoke highly of the science courses available to teachers of primary science that have been offered by their LA. The views of those science specialists, who were active within their science specialist LA networks, were that the CPD material they had received was very similar to that which they had received at the NSLC and that they thought both were very good[8]. What was most noticeable was that those teachers who were most enthusiastic about the NSLC CPD were those who had no LA support. Those teachers in schools with little, if any, LA CPD for science specialists spoke enthusiastically (and in some cases at great length) about the material they had either received from their visit to the NSLC, as part of the full CPD group, or the websites that they had been directed to for the first time whilst there which they found to be rich in terms of teaching materials. While we have not received details of the CPD material, and so cannot verify the details of the claims made by science specialists, we have, for example, heard glowing praise from a number of teachers about the material available for delivering a practical task involving the making and testing of toothpastes and the confidence that those science specialists felt the CPD had given them.

The case-study lessons observed across all three groups tended, generally speaking, to be well planned and involved ample opportunity for pupils not only to do practical work but to think about the ideas associated with that practical task and there was clear evidence of learning. In one lesson on mass and weight (control group) pupils discovered the relationship between these for themselves as a result of a very well delivered practical lesson. However, not all lessons were as effective. In one Year 4 lesson on magnets, given by a teacher in the full CPD group, the focus remained very much about ‘doing’ with little evidence of trying to engage the pupils in thinking about what they were doing. Indeed, there was little scope for getting the pupils to do or learn more about scientific ideas because they already understood the meaning of the scientific terms ‘attract’ and ‘repel’ before the lesson started.

---

[8] Some cross-over of provision is likely here as the same experts deliver CPD regionally and at the NSLC.

6.2. Post-test 1 September 2013

6.2.1. Teacher test summary

Of the 84 schools still involved in the project at post-test 1, 56 specialist tests and 44 colleague tests were received. Because tests could only be used if the same person had returned their papers at the baseline stage, the final analysable sample was 51 science specialists and 41 colleagues (92 total). This is shown in Table 5, where the columns headed 'Baseline and Post-1' refer to the useable sample. In two cases, only one of the two papers was returned and the marks for the other paper were imputed based on their score on the one submitted[9].

Table 5. Number of returns: teacher tests

Group	Specialist Post-1	**Specialist Baseline and Post-1**	Colleague Post-1	**Colleague Baseline and Post-1**	Total Baseline and Post-1
Full	19	**16**	13	**13**	29
Partial	20	**19**	18	**16**	35
Control	17	**16**	13	**12**	28
Total	56	**51**	44	**41**	92

As in the Baseline, teachers rated their confidence in their answers to each question they tackled, from 0 to 10. Again, only those teachers who had rated at least 14 out of the 29 questions (at both baseline and post-test) have been included in the following data (Table 6).

Table 6. Number of teacher tests analysed for confidence ratings

Group	Specialist	Colleague
Full	14	12
Partial	17	16
Control	16	11
Total	47	39

[9] A Pearson Correlation between Paper A and Paper B gave a significant result at the 0.01 level of 0.791

6.2.1.1. Test performance

The statistics reported in this section in square brackets are the results of one-way between-groups analyses of covariance, conducted to assess the effectiveness of the two CPD courses on specialists' and colleagues' subject knowledge. The relevant group's performance on a comparable test at baseline was used as the covariate in the analysis, enabling adjustment for pre-intervention scores when comparing post-intervention performances.

The maximum possible mark for Papers A and B combined on the teacher tests was 150 at both baseline and post-test 1. The specialists' average marks at post-test 1 were 87.3 for the full CPD group (58%), 95.9 for the partial CPD group (64%) and 94.3 for the control group (63%) (see Table 7). Taking into account the performance of teachers at baseline, there was **no statistically significant difference** in the change in scores between the different groups [$F(2, 47)=0.269$, $p=0.765$]. The test results showed that the depth of conceptual knowledge and understanding varied hugely across the specialists. The range of scores was consistently wide between stages and across groups, from a low of 33 at post-test 1 to a high of 144 (6 short of full marks).

The colleague average test scores were very similar to those for the specialists and showed a similar spread of performance (from 41 to 139 at post-test, see Table 8). Once again, there was **no statistically significant difference** in the scores between the different groups, allowing for any differences at pre-test [$F(2, 37)=0.042$, $p=0.959$].

When all trial groups were compared, there was **no statistically significant difference** between the specialists or colleagues on their change in performance between baseline and post-test 1 on any of the four Sc categories.

There was no statistically significant difference in the confidence ratings between the three groups of specialists [$F(2, 43)=0.174$, $p=0.841$] or colleagues [$F(2, 35)=0.449$, $p=0.642$]. Moreover, univariate analysis of variance showed no statistically significant difference between the specialist and colleague within any of the three treatment groups (full CPD group [$F(1, 23)=0.261$, $p=0.614$], partial CPD group [$F(1, 30)=0.083$, $p=0.775$] or control group [$F(1, 24)=0.311$, $p=0.582$]).

Table 7. Specialist test performance - baseline and post-test 1

		Baseline				Post-test 1			
	No of tests	Minimum	Maximum	Mean	Std dev	Minimum	Maximum	Mean	Std dev
Full CPD group	16	56	113	84.3	20.3	46	124	87.3	23.7
Partial CPD group	19	26	129	90.8	25.0	33	144	95.9	27.5
Control group	16	61	127	92.3	19.8	61	120	94.3	15.5

Table 8. Colleague test performance - baseline and post-test 1

		Baseline				Post-test 1			
	No of tests	Minimum	Maximum	Mean	Std dev	Minimum	Maximum	Mean	Std dev
Full CPD group	13	29	111	84.0	25.4	43	136	89.0	25.7
Partial CPD group	16	44	120	95.1	23.4	41	139	96.3	25.9
Control group	12	61	114	91.9	17.0	61	136	93.8	26.5

6.2.2. Pupil test summary

The full breakdown of descriptive statistics for the pupil baseline and post-test 1 results by treatment group is provided in Table 9 below.

Table 9. Overview of pupil post-test scores by year group and treatment

Treatment	Year Group	Number of pupils	Mean (Baseline)	Mean (Post-test)
	Year 3	172	31.02	23.78
Full CPD	Year 4	159	20.79	26.01
(18 schools)	Year 5	141	15.55	22.83
	Year 3	113	29.69	25.15
Partial CPD	Year 4	162	21.56	26.06
(25 schools)	Year 5	202	15.05	20.35
	Year 3	174	30.79	23.30
Control	Year 4	133	20.70	26.31
(21 schools)	Year 5	168	17.75	23.58

The Year 3 test means have decreased when compared to the baseline means, but this was expected because these pupils were asked to take a more difficult (Year 4) test as the post-test. All the test scores were standardised (turned into z-scores) in order to facilitate the comparison of baseline and post-tests.

After adjustment for baseline test score, the multi-level modelling was carried out and there were found to be **no significant differences in post-test scores between the treatment groups,** i.e. the full CPD group's scores were not statistically significantly higher than those of the other trial groups.

Year groups were also used as a variable and, again, **no statistically significant differences were found**. When gender was used as a variable, however, it was found that girls performed better than boys and that the difference was **statistically significant** ($p<0.001$).

6.2.3. Pupil attitudes summary

Pupils were asked to complete the same 'Attitudes to Science' questionnaire at both baseline and post-test stages. In the post-test 1 stage 1,219 completed questionnaires were returned (for both stages, questionnaires with five or more missing responses out of 44 were excluded). The responses to each question were analysed comparing baseline and post-test answers across the three treatment groups. Findings that were statistically significant are outlined below.

The first section of the pupil questionnaire contained a set of statements about being in school. Pupils were asked to rate a series of statements on a five-point scale from ‘Like it a lot’ through to ‘Hate it’. For this section, there were very few differences across the three treatment groups. Slightly larger proportions of the full CPD group pupils, by the time of post-test 1, however, liked doing science (up from 74% to 82%), and this change in scores was **statistically significantly** higher than those of the other trial groups (Kruskall-Wallis Test, $p<0.001$; and Mann-Whitney Test, $p<0.001$, for both the full CPD group compared with the partial CPD group and for the full groups compared with the control group.

The second set of pupil attitude questions was more specifically about science and science-related activities. Again, for most of these questions there was little difference in any changes of pupil opinions across the three groups, but for one question, the improvement in scores was statistically significant as regards the full treatment group in relation to the control group (but not the partial treatment group). The change in scores for the full CDP group on levels of agreement with the statement ‘Science in school is just too difficult’ was **statistically significantly** lower than that of the control group C (Kruskall-Wallis Test, $p<0.001$; and Mann-Whitney Test, $p<0.001$, for both the full CPD group compared with the partial CPD group and for the full groups compared with the control group

There were also some significant findings from the third set of questions, on ‘What about you and science’. A larger proportion of pupils in the full CPD group agreed that ‘Science makes me think’, increasing from 68% to 83% , a change in scores for this item that was **statistically significantly** higher than that of the control group (Kruskall-Wallis Test, $p<0.001$; and Mann-Whitney Test, $p<0.001$). Smaller proportions of pupils, across all three groups, felt that ‘Science is more for boys’ and that ‘You have to be clever to be a scientist’, indicating the development of less stereotypical views about the characteristics of scientists. In the case of the 'Science is more for boys’ item the change in scores (indicating a less stereotypical view) between the full CPD group was **statistically significantly** higher than that of both other groups (Mann-Whitney Test $p<0.001$ in both cases). The differences in changes in attitude scores about the statement 'You have to be clever to be a scientist' were not statistically significant, suggesting that this 'improved' attitude was present at similar levels across all three groups.

In the final section of the questionnaire pupils were asked to use the five-point agree-disagree scale in response to a set of statements about the contribution of science to society. There was slightly more (positive) variation in the attitudes towards science in society on the part of the full CPD group compared to the other two groups, though the differences were not large and **not statistically significant**.

In summary, the findings from the pupil attitude questionnaire at post-test 1 stage suggested that for a number of questions there was evidence of improved attitudes towards science across all three groups, although sometimes the largest proportional improvements were in the full treatment group.

7. Post-test 2 results

7.1. Sample composition

Table 10 shows the composition of the sample at randomisation (96 schools) compared to the sample contacted for post-test 2 (80 schools – shown in brackets).

To examine change over the two years of the evaluation, only those pupils, teachers and schools that correctly returned data at both baseline and post-test 2 have been included in the analysis[10]. Figures for post-test 1 to post-test 2 have been calculated but showed **no statistical significance** (details are provided in Table 19 and Table 20 (page 35), and Table 24 in Section 9.2). As a consequence the baseline data has been re-analysed to include only those who also supplied a test or questionnaire at post-test 2.

The exception to this is the pupil attitude data. There were significant shifts on some statements across the three stages of the research, and to allow comparison whilst ensuring consistency in analysis, the sample for this set of data is those pupils who returned the survey at all three collection points.

It is important that the three treatment groups are closely matched by key characteristics to minimise the risk of findings being skewed by effects unrelated to the intervention (see Table 10). Attainment at KS2 and percentage of pupils eligible for FSM were the two main factors used to match the triplets of schools before randomisation, hence these were very similar across the three groups initially and remained so at post-test 2 despite attrition. In terms of attainment, the average proportion of pupils achieving Level 4 or above in KS2 English and mathematics over 2009-2011 was 76% of schools in both the original and final samples, very similar to the national average of 75% for state schools in England. Average eligibility for FSM across the baseline and post-test 2 schools was 23% versus a national average of 19% (for 2011). The schools in post-test 2 were also closely comparable to the baseline sample on average school size. An examination of pre-baseline Ofsted performance showed that the profile of post-test 2 schools was similar to that of the original sample of 96 schools, indicating no bias towards low or high rated schools in terms of drop-out.

Attempts were also made to spread the year groups of the classes involved in the study evenly. Although this proved less easy to control owing to last-minute changes of science specialist and class taught, the treatment groups were reasonably balanced. Over the two years, attrition was more marked in the control group Year 3 and Year 5 in the full CPD group. Comparing all the year groups, attrition was greatest among Year 3 classes but there was no indication of why this might be.

[10] Pupil and teacher test data from post-test 1 to post-test 2 has been analysed, but no significant differences were found.

Elements of the CPD were delivered regionally, and recruitment was spread across the three regions. It was less successful in Yorkshire and the Humber, so that initially the sample size here was much lower than in the other two regions (22 compared with 37 in each of the others). However, attrition across the two years was much higher in the North West and London, leading to the final figures being slightly less uneven (20 in Yorkshire and the Humber versus 30 in each of the other two regions).

Overall, Table 10 demonstrates that, despite attrition, the sample composition remains well-matched across the three treatment groups and there is no obvious bias in terms of dropout rates across the school characteristics shown.

Table 10. Sample composition: initial with final numbers in parentheses

Schools N=96 (final N=80)	Full CPD group	Partial CPD group	Control group	Total
Total	32 (23)	32 (30)	32 (27)	96 (80)
School characteristics				
KS2 Level 4+ Eng/Maths[1]	75% (75%)	76% (77%)	76% (75%)	76% (76%)
FSM	23 (22)	24 (23)	23 (23)	23 (23)
School roll (average)	298 (323)	361 (364)[2]	338 (333)	333 (340)
Year Group in study[3]				
Year 3	13 (10)	11 (11)	17 (11)	41 (32)
Year 4	11 (8)	13 (12)	11 (11)	35 (31)
Year 5	11 (6)	11 (10)	9 (9)	31 (25)
Region				
North West	13 (8)	14 (14)	10 (8)	37 (30)
Yorkshire and Humber	8 (6)	5 (5)	9 (9)	22 (20)
London	11 (9)	13 (11)	13 (10)	37 (30)

[1] Average 2009 to 2011

[2] Includes a 4-16 academy

[3] Because of mixed year groups the total is greater than the number of schools

The number of schools still active at the end of the trial (i.e. those returning any data in the final stage) was 76 out of the original 96 schools, an attrition rate of 21%. The participating schools were distributed as follows:

- 23 in the full CPD group
- 27 in the partial CPD group
- 26 in the control group.

This represents some levelling out since post-test 1.

It should be stressed that, at each of the three data collection points, some schools returned incomplete data, and did not always provide the missing set of papers when followed up. Sometimes this was because completed papers had been lost in the school, but often no reason, despite repeated requests from the evaluators, was forthcoming. In post-test 2, all 76 returned pupil tests and attitude surveys. Fewer teacher tests were returned, but in several cases this was because the teacher was no longer at the school or unable to participate for other valid reasons. Overall, the response rate for specialists was 75% (43 out of 57) dispatched, and for colleagues it was 54% (35 out of 65 dispatched). However, some teachers had valid reasons for non-return (e.g. left the school, sickness) as shown in Table 11. Two conclusions can be drawn: firstly, pupil tests were more likely to be returned than teacher tests; and secondly, specialists' own papers were more likely to be returned than those of their colleagues.

Table 11. Reasons for non-return

Reason	**Specialist**	**Colleague**
Not sent	**23**	**14**
Left the school	17	5
No previous completions	6	9
Sent not returned	**14**	**30**
No response	12	11
No time/not willing	1	6
Justifiable reason:		
Left the school	1	8
Retired	-	2
Maternity	-	1
On leave/not in school	-	2

7.1.1. Teacher sample

For the analysis of baseline and post-test 2, tests were only included if the same individual had returned their papers at both stages. Reason for non-return at baseline included loss in the postal system and late completion (after the CPD had begun). This meant six post-test 2 returns could not be used. The final analysable sample was 40 science specialists and 32 colleagues (72 total). This is shown in the columns headed 'Baseline and Post-2' in Table 12. A power analysis shows that, with these sample sizes, the Effect Size (ES) would have to be over 0.90 to be detected at 80% power, or over 0.65 if A/B were combined. For an educational intervention, these are large ESs. In two cases, only one of the two papers was returned and the marks for the other paper were imputed based on their score on the one submitted[11].

[11] A Pearson Correlation between Paper A and Paper B gave a significant result at the 0.01 level of .698

Table 12. Number of returns: teacher tests

Group	Specialist Post-2	Specialist Baseline and Post-2	Colleague Post-2	Colleague Baseline and Post-2	Total Baseline and Post-2
Full CPD	15	12	12	12	24
Partial CPD	15	15	11	9	24
Control	13	13	12	11	24
Total	43	40	35	32	72

Under the criteria explained in Section 5.2, 70 teachers qualified for inclusion at post-test 2 alone, reducing to 60 once those who had not made a valid return at baseline were excluded. The split by treatment group is shown in Table 13.

Table 13. Number of teacher tests analysed for confidence ratings

	Specialist		Colleague	
Group	Post-2 returns	Baseline and Post-2	Post-2 returns	Baseline and Post-2
Full CPD	12	11	11	6
Partial CPD	14	14	9	7
Control	13	13	11	9
Total	39	38	31	22

7.1.2. Pupil sample

Science specialists in each of the three groups were asked to administer the pupil tests and attitude surveys to the class that they had originally been teaching as part of this study (for most of them at post-test 2, this was their class from the previous year). The numbers of pupils returning tests at baseline and post-test 2 are shown in Table 14 below. The final column shows the number that could be used in the analysis for this report (i.e. pupils who had completed tests at both stages).

Table 14. Number of returns: pupil tests

Group	Baseline		Post-2		Baseline and Post-2	
	Pupils	Schools	Pupils	Schools	Pupils	Schools
Full CPD	594	26	491	23	471	23
Partial CPD	717	31	584	30	550	27
Control	669	28	582	27	561	26
All	1,980	85	1,657	80	1,582	76

The numbers of pupil tests returned and response rates by treatment group are shown in Table 15. The response rates were slightly lower in the partial CPD group than the other two groups (mainly a reflection of school-level attrition) and reached 81% of pupils overall.

Table 15. Number of pupil tests returned by treatment group (post-test 2)

	Full CPD		Partial CPD		Control		Totals	
	Pupils	Schools	Pupils	Schools	Pupils	Schools	Pupils	Schools
Tests returned	491	23	584	30	582	27	1,657	80
Expected number of returns	595	26	756	31	691	27	2,042	84
Response rate	83%		77%		84%		81%	

The numbers of pupil baseline and post-test 2 pupil attitude questionnaires are shown in Table 16. Also shown is the number of pupils analysed across the study – here, for reasons of continuity, only those who responded at baseline, post-test 1 *and* post-test 2 have been analysed (see Section 7.1 for explanation). This gave a total of 1,782 valid baseline returns, 1,219 valid post-test 1 returns and 1,488 valid post-test 2 questionnaire returns. This equates to 965 valid surveys across all three stages that could be included in the analysis.

Table 16. Number of returns: pupil attitude questionnaires

Group	Baseline		Post-1		Post-2		Baseline, Post-1 and Post-2	
	Pupils	Schools	Pupils	Schools	Pupils	Schools	Pupils	Schools
Full CPD	545	23	404	18	419	20	270	16
Partial CPD	665	29	543	25	532	26	398	24
Control	572	25	426	21	537	26	297	20
All	1,782	77	1373	64	1,488	72	965	60

7.2. Statistical power

Given the changing pupil sample size we have continued to monitor the statistical power of the analysis and an updated power calculation has been run using the latest figures for the sample sizes. This shows that there are still sufficient schools and pupils in each treatment group for the multi-level modelling analysis to continue as planned. [Note: multi-level modelling was not carried out on teacher data because of the smaller sample size, and therefore no power analysis was needed].

Assuming pre/post correlation = 0.70

Intra-class correlation (ICC) = 0.125

Power = 0.80

Minimum detectable effect size (MDES) = +0.24 / +0.25 between the different treatment groups.

8. Quantitative data: Teacher tests and confidence ratings

This section reports the statistical analyses from teachers' subject knowledge tests and their confidence about answering the questions in those tests. The statistics reported in this section in square brackets are the results of one-way between-groups analyses of covariance, conducted to assess the effectiveness of the two CPD courses on specialist' and colleagues' subject knowledge. The relevant group's performance on a comparable test at baseline was used as the covariate in the analysis, enabling adjustment for pre-intervention scores when comparing post-intervention performances.

8.1. Test performance

As shown in Table 17 (page 35), the specialists' average marks at post-test 2 were 96.8 for the full CPD group (64%), 99.5 for the partial CPD group (66%) and 95.1 for the control group (63%). Taking into account the performance of teachers at baseline, there was **no statistically significant difference** in post-test scores between the different groups [F(2,36)=0.484, p=0.620]. The test results showed that the depth of science knowledge and understanding varied hugely across the specialists. The range of scores was consistently wide between stages and across groups. In post-test 2, scores ranged from 63 to134 out of 150. Figure 3 shows the spread of marks for all specialists combined.

One risk when running analyses with small sample sizes is that a non-significant result may be due to insufficient power[12]. The specialist analysis was repeated combining the two intervention groups to increase the sample size. This also failed to reach statistical significance [F(1,37)=0.699, p=0.408]. However, the combined sample size is still small and the security of the result still suffers from lack of power, i.e. we did not have a large enough sample to guarantee finding an effect if there was one.

The colleague average test scores were very similar to those for the specialists (Table 18) and showed an even wider spread of performance (from 47 to 142 at post-test 2, see Figure 4). There was no statistically significant difference in the scores between the three treatment groups (allowing for any differences at baseline) [F(2,28)=0.551, p=0.583]. As with the specialists analysis, the intervention groups were combined to increase the sample size, and there was still no significant different in performance [F(1,29)=0.964, p=0.334].

The test scores of specialists and colleagues were compared with each other in each of the three treatment groups, to explore whether science specialists had significantly more (or less) science knowledge than their colleagues. **No significant differences** were found [Full CPD

[12] See Appendix B for further explanation

group: F(1,21)=0.953, p=0.340; partial CPD group: F(1,21)=0.072, p=0.791; control group: F(1,21)=1.491, p=0.236].

Figure 3. Spread of scores of the teacher tests: specialists (all those included in baseline/post-test 2 analysis (N=40))

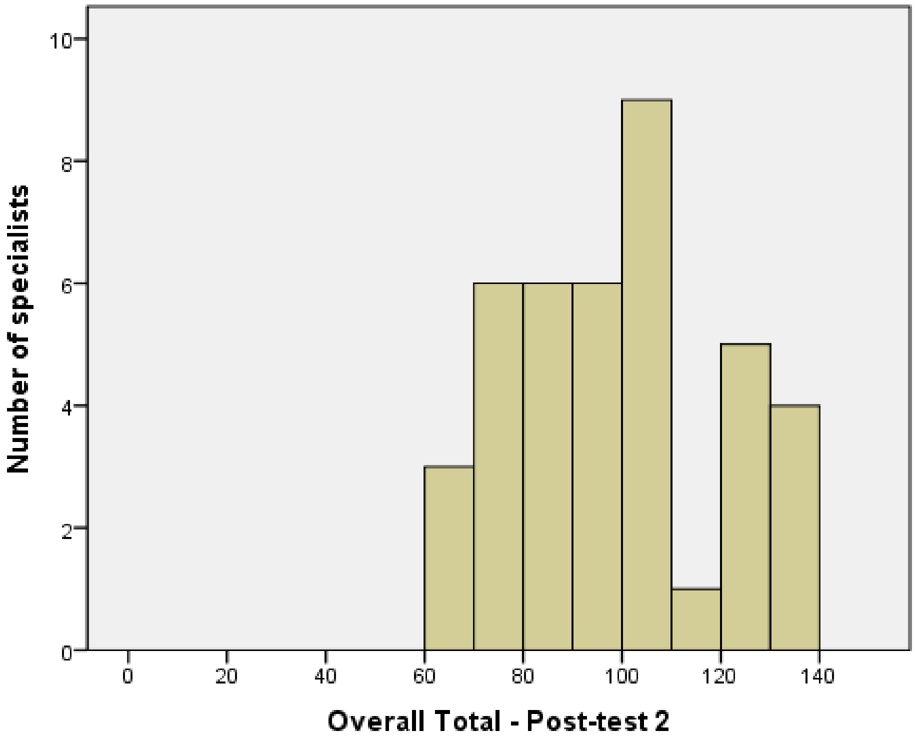


Figure 4. Spread of scores of the teacher tests: teacher colleagues (all those included in baseline/post-test 2 analysis (N=32))

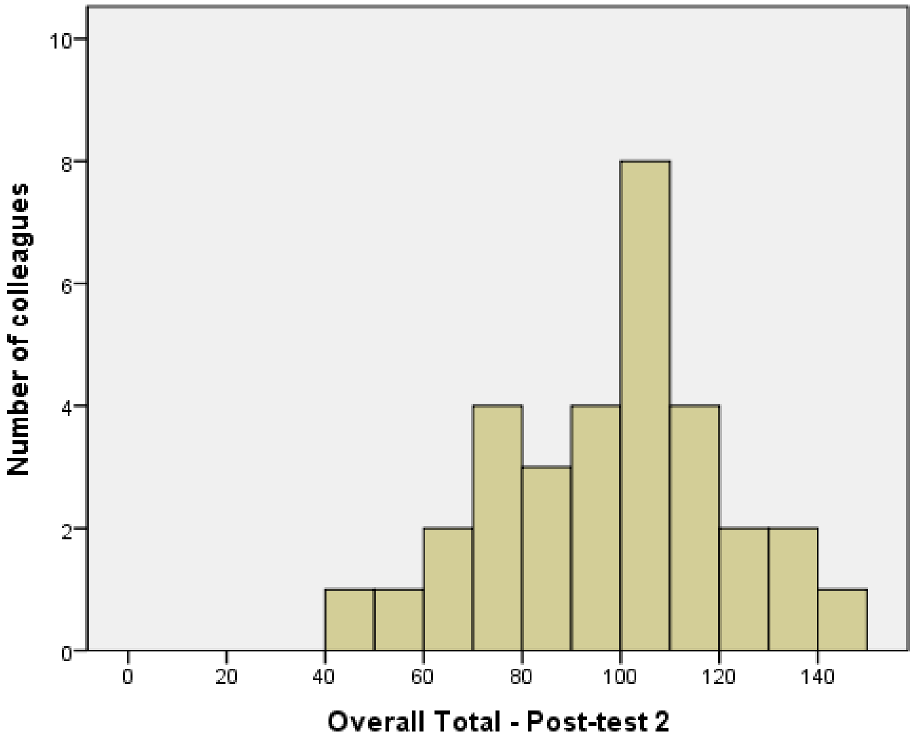


Table 17. Specialist test performance - baseline and post-test 2

		Baseline				Post-test 2			
	No of tests	Minimum	Maximum	Mean	Std dev	Minimum	Maximum	Mean	Std dev
Full CPD group	12	56	113	81.0	19.1	71	120	96.8	17.4
Partial CPD group	15	26	128	89.3	29.0	63	134	99.5	27.6
Control group	13	61	127	86.4	21.1	73	121	95.1	12.9

Table 18. Colleague test performance - baseline and post-test 2

		Baseline				Post-test 2			
	No of tests	Minimum	Maximum	Mean	Std dev	Minimum	Maximum	Mean	Std dev
Full CPD group	12	29	143	86.9	31.4	47	142	92.8	27.7
Partial CPD group	9	44	126	90.4	26.6	63	125	98.0	22.6
Control group	11	34	114	84.0	23.9	77	137	100.6	19.1

Table 19. Specialist test performance - post-test 1 and post-test 2

		Post-test 1				Post-test 2			
	No of tests	Minimum	Maximum	Mean	Std dev	Minimum	Maximum	Mean	Std dev
Full CPD group	15	46	124	83.5	24.7	64	141	98.6	21.6
Partial CPD group	14	33	132	94.1	31.4	63	134	101.8	27.2
Control group	12	60	117	91.2	18.2	73	121	94.0	12.8

Table 20. Colleague test performance - post-test 1 and post-test 2

		Post-test 1				Post-test 2			
	No of tests	Minimum	Maximum	Mean	Std dev	Minimum	Maximum	Mean	Std dev
Full CPD group	10	57	136	92.6	24.3	47	123	87.5	25.3
Partial CPD group	9	41	118	89.4	23.4	63	125	98.4	20.6
Control group	10	61	136	92.6	27.4	77	137	100.8	20.2

Directionally, the scores have all increased, but this may reflect a difference in content of the baseline and the post-test papers rather than any absolute improvement in teacher performance since the scores have increased across all three treatments – including the control.

It is important to note that specialists' gains were directionally higher among the two treatment groups than the control group (Table 17). Gains averaged +15.8 in the full CPD group, +10.2 in the partial CPD group, and +8.7 in the control group C. These comparisons were greatly underpowered, so they did not approach significance. No such trend was seen for colleagues.

8.1.1. Analysis by science curriculum categories

The questions have been analysed by the four main categories of the primary science curriculum (Sc1 – 4) using the distribution of marks in the original mark scheme. The four categories are:

Sc1 – scientific inquiry

Sc2 – life processes and living things

Sc3 – materials and their properties

Sc4 – physical processes

All trial groups were analysed separately to explore whether the performance in different Sc categories was affected by the treatment group teachers were in. There were no statistically significant differences for either specialists [Sc1 $F(2,36)=0.508$, $p=0.606$; Sc2 $F(2,36)=1.783$, $p=0.183$; Sc3 $F(2,36)=0.705$, $p=0.501$; Sc4 $F(2,36)=0.100$, $p=0.905$] or colleagues [Sc1$F(2,26)=0.937$, $p=0.405$; Sc2 $F(2,26)=0.047$, $p=0.955$; Sc3 $F(2,26)=0.963$, $p=0.395$; Sc4 $F(2,26)=0.892$, $p=0.422$].

Table 21. Teacher test performance on different Sc categories

70 teachers	Mean percentage score	Standard deviation
Sc1	64.8	19.06
Sc2	73.1	13.49
Sc3	56.2	17.09
Sc4	65.0	17.92

Leaving aside which treatment group they were in, teacher performance was weaker in some Sc categories than others. However, the pattern of these variations was different compared with the previous stages of the study. In both the baseline and post-test 1, significantly more Sc1 questions were correct than Sc2, and teachers performed significantly better in both Sc1 and Sc2 than Sc3 and Sc4. On post-test 2, the strongest performance was on Sc2 (73% average mark). The marks for Sc4 (65%) were on a par with Sc1 (65%), and performance was weakest on Sc3 (56%). Since there was no significant difference between the performance of specialists and colleagues in any of the four Sc categories, the two samples were combined to allow this analysis.

8.2. Confidence

As has been stated in the methodology, to make the analysis as robust as possible and avoid potential bias, only those teachers who had given a rating for at least 14 of the questions at baseline and 14 at post-test 2 have been included in the following analyses. The resulting sample sizes were 38 specialists and 22 colleagues.

The specialists' mean confidence rating across the two papers was 5.2, although the range was very large – the most confident specialist had a mean of 8.5, and the least confident had a mean of 1.8 (which included rating eight questions they answered as 0, "a total lack of confidence"). Figure 5 shows the spread of specialists' confidence ratings, averaged across the whole test.

The mean for colleagues was 5.1, ranging from 7.5 down to 0.2 (including a confidence rating of 0 on 22 questions). The spread is shown in Figure 6.

Figure 5. Specialists' test confidence ratings (N=38)

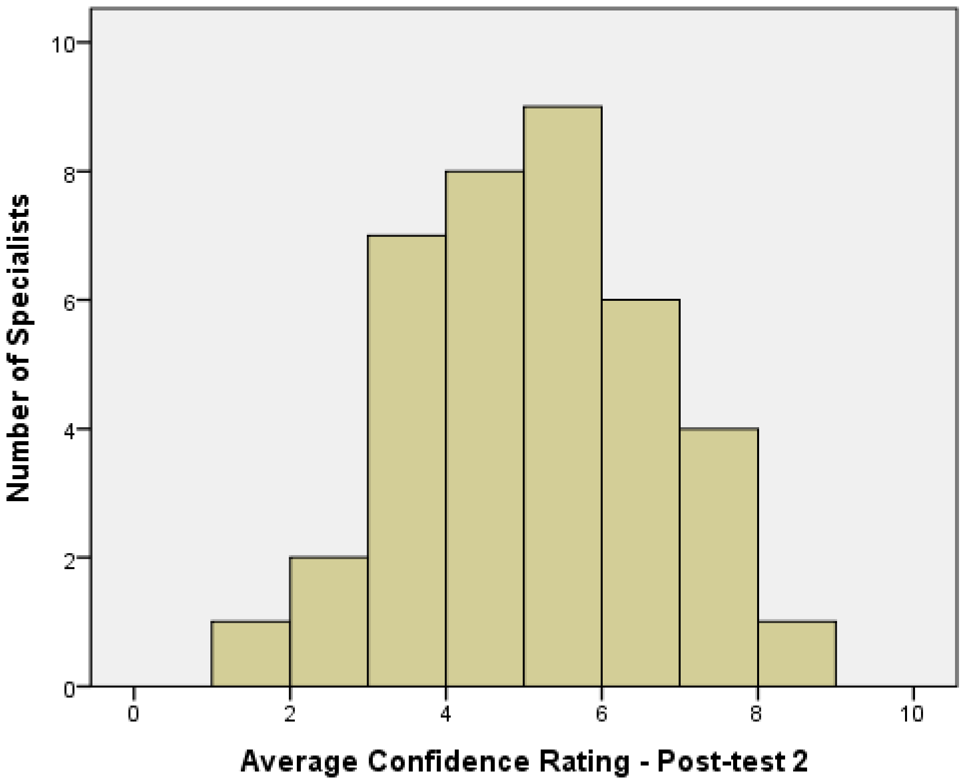
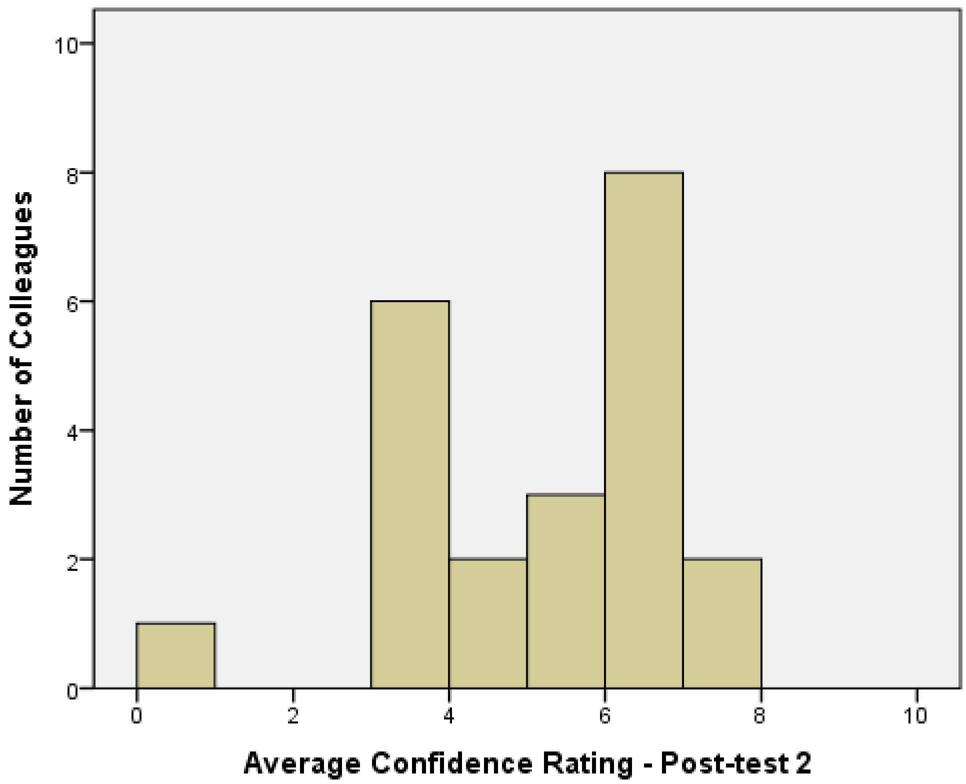


Figure 6. Colleagues' test confidence ratings (N=22)



There was **no statistically significant difference** in the average confidence ratings at post-test 2 (taking into account the average ratings at baseline) between the three groups of specialists (Table 22). The same was true for the colleagues (Table 23).

Table 22. Specialist confidence ratings - baseline and post-test 2

Group	Baseline	Post-2	No of tests	F-value	p-value
Full CPD	4.81	4.85	11		
Partial CPD	4.89	5.74	14		
Control	4.24	4.81	13		
ANCOVA				1.453	0.248

Table 23. Colleague confidence ratings - baseline and post-test 2

Group	Baseline	Post-2	No of tests	F-value	p-value
Full CPD	5.59	5.86	6		
Partial CPD	3.42	4.13	7		
Control	4.94	5.32	9		
ANCOVA				0.059	0.943

[Note: For each question, teachers rated their answers from 10=extremely confident in all your answers to 0=total lack of confidence in all your answers to this question].

Univariate analysis of variance showed there was **no statistically significant difference** between the specialists and colleagues within the full CPD group [F(1, 14)=0.664, p=0.429], partial CPD group [F(1, 18)=0.426, p=0.522] or control group [F(1, 19)=0.284, p=0.600)].

Overall, there was a medium, positive correlation between test performance and confidence rating ($r=0.56$, $N=60$, $p<0.01$[13]). That is, the general pattern was for teachers to be more confident on the questions they answered correctly, and less confident about those they did poorly on.

There were insufficient respondents in the sub-samples to look for any evidence of patterns between the different roles or treatment groups.

Despite the comparatively low levels of achievement, there is some sense that teachers' confidence levels were even lower than might be anticipated by their performance: the teacher with the lowest confidence rating overall (0.2, i.e. no confidence at all in the vast majority of answers) actually got a mark of 63 (equivalent to 42%).

[13] Pearson product-moment correlation coefficient

9. Quantitative data: Pupil tests and attitude measures

9.1. Introduction

A key research aim from the start of this evaluation was to provide evidence about the impact of the CPD course on pupils in terms of their science-specific attitudes, self-concepts and achievement. Pupil tests were used to look at achievement and a pupil survey was used to look at attitudes, both in relation to science, and in relation to school (and self-concepts) more generally. This section of the report outlines the key evidence obtained in relation to these types of impacts upon pupils.

9.2. Analyses of pupil test scores

We worked with the central hypothesis that the achievement of pupils in the tests would improve more for pupils in the treatment groups than those in the control group. In order to identify whether there were any differences between the three treatment groups, at each of the three stages, pupil test scores were standardised so that we could compare scores across the three year groups. A summary of all the modelling analyses across the three time points is provided in Table 24.

Table 24. Overview of analyses of pupil test scores (all year groups combined) across the three stages of the main study

Baseline	A one-way between-groups analysis of variance showed **no significant differences (**p=<0.05**) between treatment groups**. This meant that from the beginning of the randomised controlled trial we had pupil groups of broadly similar ability.
Baseline to post-test 1	At the **post-test 1** stage, after adjustment for baseline test score, the multi-level modelling was carried out and there were found to be **no statistically significant differences in post-test 1 scores between the treatment groups,** i.e. the full CPD group's scores were not statistically significantly higher than those for the other groups.
Post-test 1 to post-test 2	A one-way between-groups analysis of variance showed **no significant differences in pupil test outcomes between the treatment groups** in post-test 2 compared to post-test 1.
Baseline to post-test 2	After adjustment for baseline test score, the multi-level modelling was carried out and there were found to be **no statistically significant differences in post-test 2 scores between the treatment groups**: the scores of the full CPD group were not statistically significantly higher than those of the other groups.

The main analyses for this final evaluation report have been from baseline to post-test 2 and are designed to examine whether there has been any [statistically significant] benefit for pupils' achievement over the two-year period from when the science CPD commenced. The analyses at this stage were based on 1,582 pupils from 76 schools who had completed the tests at both baseline and post-test 2 stages (see Table 14 on page 31) – they are therefore *individually matched* test scores.

After adjustment for baseline test score, the multi-level modelling (MLM) was carried out and there were found to be **no statistically significant differences** in post-test 2 scores between the treatment groups.

The fact that the RCT had three treatment groups provided another opportunity to look at the findings using MLM. The two CPD groups were combined and then their results were compared with those of the control group. Again, there were found to be **no statistically significant differences** in scores between these two groups**.** (For full details of the MLM outcomes, see Appendix D.**)**

Following the finding that there were no statistically significant differences between the pupil test scores of the treatment groups, we went on to calculate effect sizes: these help to 'quantify' the differences between any two groups. After adjustment for the covariates for post-test 2 pupil scores, the effect size found for the full CPD group versus the control was 0.02, for the partial CPD group versus the control, -0.10 and for both CPD groups combined versus the control was -0.05.

9.3. Analyses of pupil attitude questionnaires

As noted previously (see Section 7), these analyses of pupil attitude data are based on the views of pupils who returned the survey data at *all three collection points*. The three survey points were included because there were some potentially important shifts in pupils' views on some statements across the three stages of the research. The tables presented below (Table 25 to Table 36) provide information on each of the four sections of the survey. The text identifies findings which are particularly interesting and states whether or not the differences were statistically significant.

It is worth noting an analysis of the findings from the baseline attitudes survey and the post-test 1 survey revealed that attitude patterns for most questions were very similar across the three treatment groups. There were a few responses, however, which showed that full CPD group pupils' attitudes had improved more than those of pupils in the other two groups. One example of this was that a larger proportion of pupils in the full CPD agreed that 'Science makes me think', increasing from 68% to 83% in this group.

At the baseline stage, 1,782 valid completed questionnaires were returned, in the post-test 1 stage the equivalent figure was 1,291 questionnaires, and for post-test 2 there were 1,488 completed questionnaires (for all stages, questionnaires with five or more missing responses were excluded in order to increase the clarity of the findings and to remove confusion arising from 'missing data'). Of these 965 pupils had completed a survey at each of three stages. Their responses to each question were analysed comparing baseline and post-test 1 and post-test 2 answers across the three treatment groups. Findings of interest (i.e. where there appeared to be important differences between treatment groups), across the four sections of the questionnaire, are reported below.

9.3.1. Section A - What do you think of being in school?

The first section of the pupil attitudes questionnaire contained a set of statements about being in school. Pupils were asked to rate a series of statements on a five-point scale from ‘Like it a lot’ through to ‘Hate it’. Details of responses to these statements, combining the ‘Like it a lot’ and ‘Like it a bit’ responses are provided in Table 25.

Table 25. Proportions of pupils who liked selected aspects of school 'a lot' or 'a bit': comparisons by stage and by treatment group

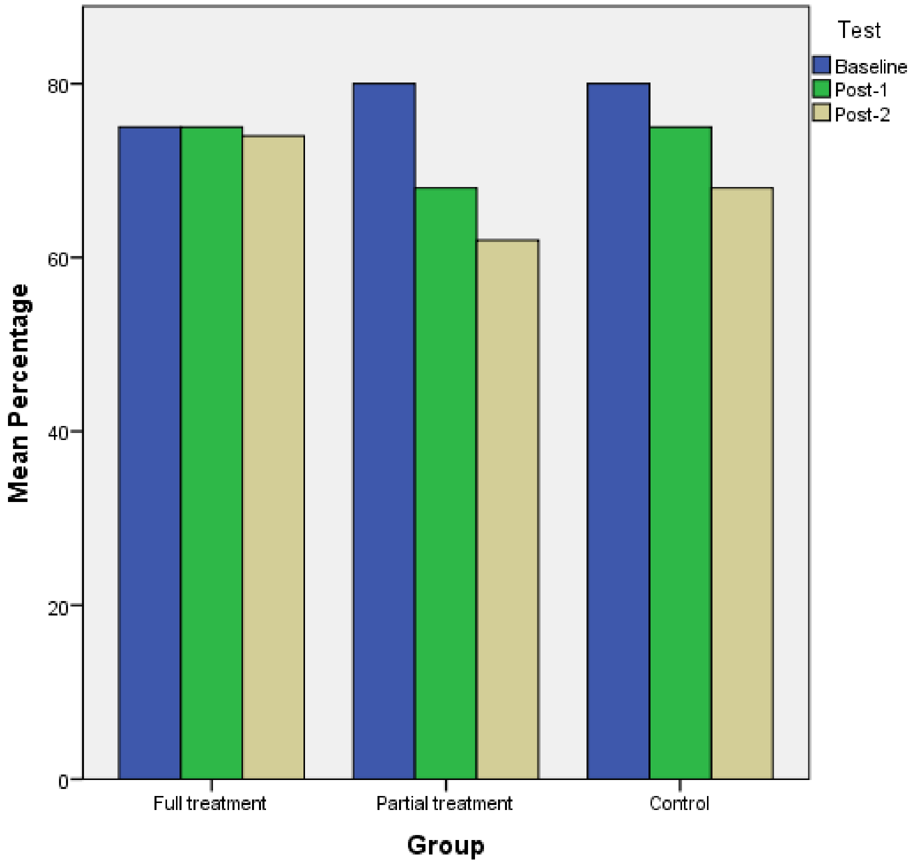
Activity	Full CPD Group (%)			Partial CPD Group B (%)			Control Group (%)		
	Base	Post-1	Post-2	Base	Post-1	Post-2	Base	Post-1	Post-2
Writing	74	74	73	74	73	69	76	82	78
Reading	83	83	75	81	81	79	88	86	81
Doing maths	75	77	78	75	76	80	79	75	81
Doing science	71	80	62	75	68	68	81	73	64
Doing science experiments	87	91	90	88	89	91	90	91	89
Using the computer	93	92	92	94	94	94	95	95	93
Listening to your teacher	75	75	74	80	68	62	80	75	68
Working by yourself	66	66	66	66	62	61	69	73	68
Working with your friends	88	90	88	90	90	89	90	90	92
Coming to school	69	70	70	72	65	71	72	76	71

Note: no significant differences

In general it can be seen that there were few major differences across the three treatment groups: in relation to the percentages liking certain specified activities, such as ‘using the computer’ and ‘working with your friends’, the percentages remained remarkably similar both across treatment groups and over time. There was only one item where there was an obvious difference in findings for the treatment group *vis-à-vis* the other two groups. In relation to ‘listening to your teacher’, the proportion of full CPD group pupils liking this activity remained steady at 75% or 74%, but in both other groups this proportion declined (by 18 percentage points in the case of the partial CPD and 12 percentage points in the case of the control group) (see Figure 7).

Where there were notable differences in responses we conducted tests to see if the differences were statistically significant. These tests were carried out between treatment groups, to see if there were any differences between groups at each of the baseline and post-test stages; and also over time, using the *change in scores* for any particular response item between baseline and post-test stages. The change in scores for the full CPD group, for example on the ‘listening to your teacher’ item, was **not statistically significantly different** over the two years when compared with those of the other two groups (using the Kruskall-Wallis Test).

Figure 7. Percentage liking 'listening to your teacher' - change by treatment group across test stages



With respect to ‘doing science’, the proportions of pupils in the full CPD group liking this activity has fluctuated from 71% up to 80% and then down to 62%. It could be that the impact of the teachers’ CPD was stronger after one year than it was after two years, but there will be a large number of factors at play here (including most pupils having spent the second year in another teacher’s class), and it should be noted that the proportions of pupils liking this activity in both other group also declined from baseline to post-test 2.

9.3.2. Section B - What do you think of science in school?

The second set of pupil attitude questions was more specifically about science and science-related activities. Selected findings are presented in Table 26 below.

It can be seen from Table 26 that for most of these questions, again there were few major differences in changes of pupil opinions across the three groups. For example, in relation to the item ‘choosing your own equipment’, the percentages liking this activity remained remarkably similar both across treatment groups and over time. Looking at the proportions carefully, however, there are some differences between groups which are worth looking at.

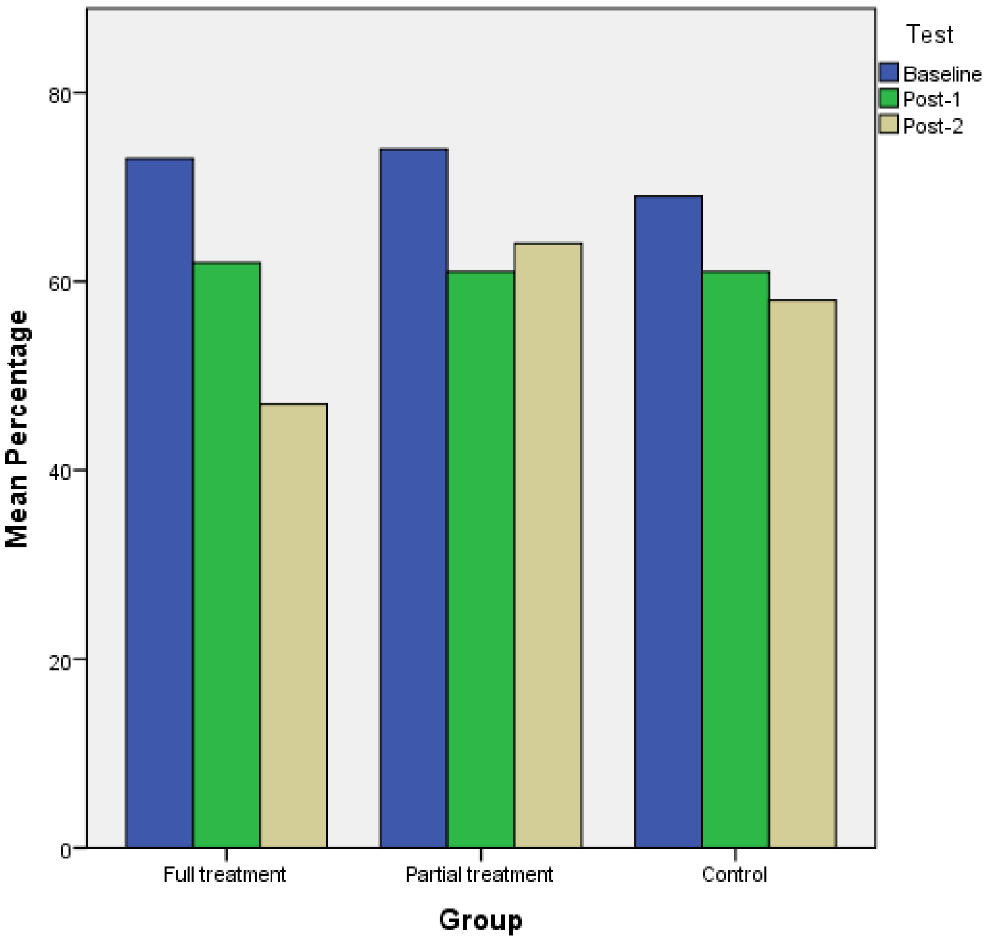
Table 26. Proportions of pupils who like various aspects of science 'a lot' or 'a bit': comparisons by stage and by treatment group

Activity	Full CPD Group (%)			Partial CPD Group (%)			Control Group (%)		
	Base	Post-1	Post-2	Base	Post-1	Post-2	Base	Post-1	Post-2
Watching the teacher do an experiment[a]	73	62	47	74	61	64	69	61	58
Working out what to do yourself	57	61	60	59	58	63	61	64	66
Teacher telling you what to do	66	61	64	69	58	61	67	65	58
Choosing your own equipment	87	85	88	88	87	87	90	89	88
Finding out what happens yourself	71	79	82	74	75	81	81	81	80
Finding out why the experiment works	70	70	67	71	67	66	73	74	67
Telling teacher what you have done	75	74	72	76	72	64	75	73	68
Telling friends what you have done	73	71	74	75	74	77	79	76	73

a: $p < 0.001$

Over the two years from the baseline survey to post-test 2, the proportions liking ‘watching the teacher do an experiment’, have declined, but the decline for the full CPD group of 26 percentage points has been more dramatic than that for either the partial CPD group (10 percentage points) or the control group (11 per cent). (This difference was found to be **statistically significant** – over the two years the change in scores for the full CPD group was statistically significantly higher than those of the other groups: Kruskall-Wallis Test, $p<0.001$). We can only speculate on the causes of this, but it does seem that pupils in the full CPD group, who have been encouraged and enjoyed doing more experiments for themselves with the CPD teacher, now do not want to just watch their teacher doing demonstrations (see also Figure 8). On the other hand, this pattern is reversed (but is nowhere near as dramatic) for the item ‘teacher telling you what to do’. In this case, over the two year period, the proportion of the full CPD group liking this activity has declined by just two percentage points, compared with eight percentage points for the partial CPD group and nine points for the control group (though for this item the difference was **not statistically significant**).

Figure 8. Percentage liking 'watching the teacher do an experiment' - change by treatment group across test stages



Another small, but possibly noteworthy, finding relates to ‘finding out what happens yourself’: in relation to this item, over two years, the proportions of both full and partial CPD groups liking this activity have increased (by 11 and seven per cent respectively), whereas the proportion in the control group has actually fallen (although only by one per cent) (although these differences were **not** found to be **statistically significant**).

This section also included four questions directly asking pupils for their views about science. The answers to these questions, suggest that there were *across the board* (i.e. across all three treatment groups) improvements in pupils’ attitudes to science.

Consideration of the time dimension in relation to particular items, however, adds one or two complexities to this more general finding. At the post-test 1 stage (the 2013 survey) there was a clear pattern of an increasing proportion of pupils in the full CPD group becoming (even) more confident or positive about science activities than pupils in the other two groups. By 2014, however, although larger proportions of the full CPD group were still more confident/more positive, the proportions in the other two groups had similarly increased. Again, we can only speculate, but this may show that the teachers' science CPD had a bigger impact on pupil attitudes after one year of the training, rather than two years after the commencement of the CPD when they had been taught by another teacher for a year in most cases. This could also reflect the fact that in the second year pupils in the full CPD group may not have had the same teacher, or that some science teachers moved to different schools taking their enthusiasm and expertise with them.

Table 27. Proportions of pupils who 'agreed' or 'agreed a bit' with selected statements about school science: comparisons by stage and by treatment group

Activity	Full CPD Group (%)			Partial CPD Group (%)			Control Group (%)		
	Base	Post-1	Post-2	Base	Post-1	Post-2	Base	Post-1	Post-2
Science in school is just too difficult	39	19	29	33	26	21	24	18	19
We do too much science at school	28	17	14	30	26	16	23	17	14
We have to do too much writing in science	48	32	45	47	40	43	42	35	41
We have to do too many experiments in science	30	17	17	29	23	16	20	15	11

Note: no significant differences

9.3.3. Section C - What about you and science?

Table 28. Proportions of pupils who 'agreed' or 'agreed a bit' with various statements about themselves and science: comparisons by stage and by treatment group

Activity	Full CPD Group (%)			Partial CPD Group (%)			Control Group (%)		
	Base	Post-1	Post-2	Base	Post-1	Post-2	Base	Post-1	Post-2
I often do science experiments at home	39	32	28	39	37	29	36	34	29
I like science more than any other school homework	53	52	35	49	41	33	50	42	35
I think that science is important	65	74	68	70	73	71	75	77	78
I should like to be given a science kit as a present	74	71	55	71	67	58	76	71	55
I like books and magazines about science	48	47	35	57	45	41	53	50	40
My father thinks that science is important	45	41	43	49	45	43	51	48	47
I like to watch science courses on TV[14]	53	59	50	57	59	56	61	66	53
Science makes me think	66	82	72	68	71	72	72	75	76
I should like to be a scientist	38	31	25	43	38	31	42	30	25
Science is more for boys	42	24	16	31	22	16	30	18	16
You have to be clever to do science	61	45	39	58	48	39	57	40	30
My mother thinks that science is important	48	44	49	51	51	45	53	52	53

The over-riding pattern in Table 28 appears to be one of the proportions of pupils (regardless of treatment group) supporting (mostly) positive statements about science activities

[14] This was the wording used in the original questionnaire (which was kept for purposes of consistency). It is likely that pupils will have interpreted ‘courses’ as ‘programmes’.

decreasing over the two years. Taking the first item, ‘I often do science experiments at home’ as an example, the percentage agreeing or agreeing a bit with this statement declined in all three treatment groups by between seven and 10 per cent over the two years. There was a similar pattern for most other items in this set of questions. There were a few exceptions, however. For example, broadly speaking, the proportions who thought that ‘science is important’ remained relatively stable for all three groups over the two-year period.

Interestingly, the biggest changes in attitudes occurred in relation to the two questions about ‘stereotypes’ in science. Across all three treatment groups there were substantial decreases in the proportions agreeing, or agreeing ‘a bit’, that ‘science is more for boys’ and ‘you have to be clever to do science’. However, the change occurred across all three groups (over the two years, the differences in changes across groups were **not statistically significant**) and so might reflect the fact that all of the schools developed a better awareness of science as they had engaged with the study.

9.3.4. Section D - What do you think of science in society?

In the final section of the questionnaire pupils expressed their views on a five-point agree-disagree scale in response to a set of statements about the contribution of science to society.

Table 29. Proportions of pupils who 'agreed' or 'agreed a bit' with selected statements about science in society: comparisons by stage and by treatment group

Activity	Full CPD Group (%)			Partial CPD Group (%)			Control Group (%)		
	Base	Post-1	Post-2	Base	Post-1	Post-2	Base	Post-1	Post-2
Science is good for everybody	72	79	72	74	72	71	77	82	75
Science helps poor people	47	39	36	43	47	40	45	44	42
Science can make my home town a better place to live	47	59	57	57	63	66	54	66	63
Lots more money should be spent on science	31	39	36	37	33	35	41	42	38
Science has made us better and safer medicines[a]	68	82	86	72	81	83	81	88	87
TV, telephones and radio have all needed science	61	74	74	64	75	77	67	82	79
Our food is safer thanks to science	50	64	63	53	62	67	60	66	71
There are lots of science in my town[15]	39	35	31	42	37	35	36	45	38
People outside school help me to understand science	44	41	41	48	47	39	46	49	40
Science makes living easier in my home town	48	54	52	47	47	52	45	58	53

a: $p < 0.05$

[15] This was the wording used in the original questionnaire (and despite the clumsiness of the phrase. it was kept for purposes of consistency). Hopefully pupils will have recognised that it was about science activities and events in their town.

There seem to be few clear patterns or trends in Table 29, suggesting that the experience of science in school can have variable impacts upon pupils' views of science in society. There are also very few instances where the opinions of the full CPD group pupils appear to have changed in a different way, or more extensively, than in the other two pupil groups.

One exception to this relates to the statement 'science has made us better and safer medicines'. Over the two years the proportions agreeing or agreeing 'a bit' with this statement have increased in all three groups, but the increase for the full CPD group (of 18 per cent) is considerably higher than the increases for the partial CPD and control groups (11 and six per cent, respectively). (This difference was found to be **statistically significant** – over the two years the change in scores for the full CPD group was statistically significantly higher than those for Groups B and C: Kruskall-Wallis Test, $p<0.05$).

10. Extension study quantitative data: pupil tests and attitude measures

10.1. Introduction

In order to be able to follow the potential impact of the CPD-trained science specialists an extension to the main study was carried out (referred to in this report as the 'extension study'). This involved sending pupil test papers and attitude surveys to 33 science specialists who took over a *new* KS2 class of pupils at the start of the academic year[16]. These schools were all volunteers from the main study. This enabled us to follow any effects upon not only the previous classes, but also the science specialists' new classes. The baseline tests and surveys for the extension schools were completed by October 2013 and the follow-up tests and surveys were completed in June 2014, so there was only one post-test (for numbers see Table 30 below).

Table 30. Extension study: school and pupil numbers from baseline to follow-up

Extension study	Pupil tests			Pupil attitude surveys	
	Number of schools	Number of pupils		Number of schools	Number of pupils
Full CPD Group	6	150	Full CPD Group	6	163
Partial CPD Group	15	357	Partial CPD Group	15	261
Control Group	12	254	Control Group	12	225
Totals	33	761	Totals	33	649

Note: Numbers of pupils refer to those individual pupils who completed the test or the attitude survey at both baseline and follow-up stages

10.1.1. Analyses of pupil test scores

The aim of the extension study was to assess whether achievement in the tests would improve more for pupils in the full and partial CPD groups than those in the control group for this smaller sample of schools. In order to identify whether there were any differences between the three treatment groups at the post-test stage, pupil test scores were standardised, so that we could compare scores across the three year groups.

The analyses of the baseline test scores were based on test papers from 761 pupils in 33 schools (response rate from schools at both baseline and post-test 2 was 100%). Detailed

[16] Not all schools were eligible to take part as some specialists were teaching in a different key stage or had moved schools.

analysis at this stage revealed that there were **no statistically significant differences** in the test results across the three treatment groups in the extension study. This meant that, in terms of test ability, at the baseline point, we had broadly similar groups of pupils across the three treatment groups.

At the post-test stage, there was found to be **a statistically significant difference** in the standardised test scores of the full CPD group compared to the partial CPD group: significance level p=0.004, effect size 0.23. There were **no statistically significant differences** between the test results for the full CPD group compared with the control, nor between the partial CPD group and the control (see Table 31 for detail). A comparison was also made of the test results for the full CPD group against the other groups combined (based on the fact that these two groups had no or low levels of CPD), and the difference was found to be **statistically significant**: significance level p=0.005, effect size 0.19 (see bottom row of Table 31).

Table 31. Analyses of all treatment group pairs for pupil test results in the extension study

		Post-adjusted			ES Confidence interval		
Condition	N	Mean	Significance	Effect size (ES)	Lower	Upper	Standard error
Partial CPD Control	357 254	-0.077 0.021	0.291	-0.1024	-0.26	0.06	0.059
Full CPD Control	150 254	0.148 0.021	0.251	0.1326	0.26	0.08	0.073
Full CPD Partial CPD	150 357	0.148 -0.077	0.004*	0.2309	0.03	0.41	0.069
Full CPD Partial CPD and Control	150 611	0.148 -0.036	0.005*	0.1902	0.01	0.36	0.069
*Significant at the p<0.05 level							

The statistically significant difference between the scores of the full CPD and partial CPD groups is interesting, but it is difficult to extrapolate why this difference was present: a significance test was carried out for the same schools in the main study and no differences were found. Thus, the difference occurred in relation to test results for pupils in the science specialists' new classes, and there were no differences in the results for the previous/old classes. There is a possibility that the effect of the CPD has become stronger, for some reason, when the science specialist has taken on a new class, one year or so on from completing the full training.

10.1.2. Analyses of pupil attitude questionnaires

In addition to the test papers, Extension Study pupils were given the 'Attitudes to Science' questionnaire that was also used in the main study. The same questionnaire was completed at both baseline and post-test stages. These analyses of pupil attitude data are based on the views of pupils who returned the survey data at *both collection points*. Findings of interest (i.e. where there appeared to be important differences between treatment groups), across the four sections of the questionnaire, are reported below.

10.1.2.1. Section A - What do you think of being in school?

The first section of the pupil attitudes questionnaire contained a set of statements about being in school. Details of responses to these statements, combining the 'Like it a lot' and 'Like it a bit' responses (from a total of five possible responses) are provided in Table 32 below.

Table 32. Proportions of pupils who liked selected aspects of school 'a lot' or 'a bit': comparisons by stage and by treatment group

Activity	Full CPD Group (%)		Partial CPD Group (%)		Control Group (%)	
	Base	Post-test	Base	Post-test	Base	Post-test
Writing	74	69	74	69	74	72
Reading	82	83	87	77	84	74
Doing maths	77	74	78	75	83	77
Doing science	82	73	79	65	85	67
Doing science experiments	93	96	92	87	90	88
Using the computer	93	93	89	91	95	89
Listening to your teacher	80	70	84	76	85	70
Working by yourself	69	61	71	67	71	65
Working with your friends	88	91	89	90	85	86
Coming to school	75	62	73	66	75	64

For many of the question items there was a similar pattern across the treatment groups and in most cases the proportion liking the selected aspects of school declined by the time of the second (post-test) survey. This was true, for example, in relation to 'writing' 'doing maths' doing science' and 'coming to school'. There were two items, however, where a different trend was evident (though **not statistically significant**):

- the proportion of pupils in the full CPD group liking reading actually increased by one per cent, whereas it dropped by 10 per cent in both of the other groups;

- there was a similar pattern in relation to ‘doing science experiments’: the proportion liking this activity in the full CPD group increased by three per cent, whereas it declined by five per cent in the partial group and by two per cent in the control group.

10.1.2.2. Section B - What do you think of science in school?

The second set of pupil attitude questions was more specifically about science and science-related activities. Responses from the Extension Study pupils are presented in Table 33 below.

Table 33. Proportions of pupils who liked various aspects of science 'a lot' or 'a bit': comparisons by stage and by treatment group

Activity	Full CPD Group(%)		Partial CPD Group(%)		Control Group(%)	
	Base	Post-test	Base	Post-test	Base	Post-test
Watching the teacher do an experiment	78	45	76	68	80	61
Working out what to do yourself	50	59	67	62	64	55
Teacher telling you what to do	69	52	69	59	71	58
Choosing your own equipment	91	91	88	90	93	88
Finding out what happens yourself	72	81	81	77	76	72
Finding out why the experiment works	77	72	72	70	79	64
Telling teacher what you have done	76	75	79	69	79	73
Telling friends what you have done	80	70	76	77	78	77

It can be seen from Table 33 that for most (but not all) of these questions, again, there is a dip in the proportion liking the activity at the second survey point across all treatment groups. But there are also a few points of interest:

- in relation to the activity ‘watching the teacher do an experiment’ there has been a decrease in the proportion liking this activity across each of the three treatment groups: but this decrease has been considerably larger, at 33 per cent for the full CPD group, than for either the partial CPD group (eight per cent) or the control group (19 per cent). This change in score for the full CPD group with respect to the other groups was **statistically significant** using $p<0.05$ as the significance level, $p=0.000$.

- the proportion of pupils in the full CPD group liking 'working out what to do for yourself' has increased from 50% to 59%, whereas this proportion has decreased in both other groups: there is a very similar pattern in relation to the item 'finding out what happens for yourself', suggesting that the science specialists' CPD has perhaps played some role in helping the pupils to work more independently. This change in score for the full CPD group with respect to the other groups was **statistically significant**, $p=0.009$.

This section also included four questions directly asking pupils for their views about science (see Table 34 below). These responses were somewhat mixed, though for three of the four items it seems that there were improvements in attitudes across the three treatment groups, and for all four items the biggest improvements in attitudes occurred in the full CPD group . For example:

- there was a considerably larger reduction in the proportion of full CPD group pupils (compared with the other two groups) who felt that 'science in school is just too difficult' (down eight per cent compared with three per cent and 0 per cent in the other two groups). This change in score for the full CPD group with respect to the other groups, however, was **not statistically significant**, $p=0.217$.
- similarly, there was a considerably larger reduction in the proportion of the full CPD group (compared with the other two groups) who felt that 'we do too much science at school' (down 11 per cent compared with six per cent and four per cent in the other two groups). This change in score for the full CPD group with respect to the other groups, however, was **not statistically significant**, though it was close to level of significance, $p=0.053$.
- the most noticeable difference was in relation to 'we have to do too much writing in science', where the proportion taking this view in the full CPD group declined by ten per cent, while the proportions in the partial and control groups increased by 11 per cent and eight per cent, respectively. This change in score for the full CPD group with respect to the other groups was **statistically significant**, $p=0.001$.
- on the fourth item, again, there was a considerably larger reduction in the proportion of the full CPD group (compared with the other two groups) who felt that 'we have to do too many experiments in science' (down 15 per cent compared with four per cent and eight per cent in the other two groups). This change in score for the full CPD group with respect to the other groups was **statistically significant**, $p=0.005$.

Overall, the responses to these four questions suggest that there has been an improvement in the attitudes towards school science for those pupils in the full CPD group and that this improvement has been larger in scale than any improvements for the other two groups.

Table 34. Proportions of pupils who 'agreed' or 'agreed a bit' with selected statements about school science: comparisons by stage and by treatment group

Activity	Full CPD Group (%)		Partial CPD Group (%)		Control Group (%)	
	Base	Post-test	Base	Post-test	Base	Post-test
Science in school is just too difficult	26	18	30	27	24	24
We do too much science at school	23	12	24	18	21	17
We have to do too much writing in science	46	36	38	49	43	51
We have to do too many experiments in science	33	18	26	22	27	19

10.1.2.3. Section C - What about you and science?

The over-riding pattern in Table 35 is one of the proportions of pupils (regardless of treatment group) supporting (mostly) positive statements about science activities *decreasing* between the first and second surveys.

Table 35. Proportions of pupils who 'agreed' or 'agreed a bit' with various statements about themselves and science: comparisons by stage and by treatment group

Activity	Full CPD Group (%)		Partial CPD Group (%)		Control Group (%)	
	Base	Post-test	Base	Post-test	Base	Post-test
I often do science experiments at home	46	42	42	35	36	26
I like science more than any other school homework	58	48	58	48	62	45
I think that science is important	81	79	78	72	80	73
I should like to be given a science kit as a present	78	72	77	66	79	72
I like books and magazines about science	56	53	56	43	60	40
My father thinks that science is important	46	45	47	41	53	41
I like to watch science courses on TV[17]	62	60	54	51	57	51
Science makes me think	78	82	76	76	75	71
I should like to be a scientist	49	42	47	37	43	34
Science is more for boys	33	20	32	26	32	21
You have to be clever to do science	64	42	54	47	62	51
My mother thinks that science is important	51	55	54	44	54	47

[17] This was the wording used in the original questionnaire (which was kept for purposes of consistency). It is likely that pupils will have interpreted ‘courses’ as ‘programmes’.

Across the twelve questions, in all three groups, there are only two responses (out of a total of 36) where the proportion taking a 'positive' view has increased – and both of these are in the full CPD group:

- the first of these is in relation to the item 'science makes me think', where the proportion in the full CPD group agreeing with the statement increases from 78 to 82 per cent, while the proportion for the partial CPD group stays the same and the proportion for the control group decreases by four per cent. This change in score for the full CPD group with respect to the other groups, however, was **not statistically significant**, p=0.091.
- the second is in relation to the item 'my mother thinks that science is important', where the proportion in the full CPD group agreeing with the statement increases from 51 to 55 per cent, while the proportions for the partial and control groups decrease by ten per cent and seven per cent, respectively. This change in score for the full CPD group with respect to the other groups, however, was **not statistically significant**, p=0.459.

10.1.2.4. Section D - What do you think of science in society?

In the final section of the questionnaire pupils expressed their views on a five-point agree-disagree scale in response to a set of statements about the contribution of science to society.

As was the case with the main study, there seem to be few clear patterns or trends in Table 36, suggesting that the experience of science in school can have variable impacts upon pupils' views of science in society. There were, however, a few instances where the opinions of pupils in the full CPD group appear to have changed in a different way, or more extensively, than in the other two pupil groups. One example of this relates to the statement 'science helps poor people'. The proportion agreeing or agreeing 'a bit' with this statement in full CPD group has increased from 49 to 58 per cent, whereas in the partial CPD group it has increased by just three per cent, from 47 per cent to 50 per cent, and in the control group it has actually decreased by ten per cent. This change in score for the full CPD group with respect to the other groups was **statistically significant**, p=0.029. Another example occurs in relation to the statement 'there [is] lots of science in our town', where the proportion of the full CPD group pupils who take a positive view increases from 33 per cent to 43 per cent, but decreases in both the other groups. This change in score for full CPD group pupils with respect to the other groups was **statistically significant**, p=0.022.

Table 36. Proportions of pupils who 'agreed' or 'agreed a bit' with selected statements about science in society: comparisons by stage and by treatment group

Activity	Full CPD Group (%)		Partial CPD Group (%)		Control Group (%)	
	Base	Post-test	Base	Post-test	Base	Post-test
Science is good for everybody	82	79	79	71	82	70
Science helps poor people	49	58	47	50	51	41
Science can make my home town a better place to live	58	72	56	62	52	58
Lots more money should be spent on science	52	50	47	38	39	33
Science has made us better and safer medicines	83	88	81	87	78	82
TV, telephones and radio have all needed science	72	83	69	79	69	72
Our food is safer thanks to science	63	76	68	66	54	61
There [is] lots of science in my town[18]	33	43	37	35	42	33
People outside school help me to understand science	52	51	48	45	48	41
Science makes living easier in my home town	55	70	55	49	44	49

[18] This was the wording used in the original questionnaire (and despite the clumsiness of the phrase. it was kept for purposes of consistency). Hopefully pupils will have recognised that it was about science activities and events in their town.

11. Qualitative data: Interview and observation data from the second round of school visits

11.1. Introduction

This section presents the results of the analysis of interview data, observation data and pupil focus group data from selected schools that participated in the randomised controlled trial.

The purpose of gathering these data was to complement the quantitative data discussed earlier. Given that much of the data is self-report data, and in order to provide as strong an evidence base as possible, data were gathered from 25 schools, representing a 33% sample of the schools that participated in the RCT. Whilst there were initially 30 case study schools five subsequently declined to take part in the second round of visits despite repeated requests to the head teachers and the science specialists to reconsider. As with the first round of visits, the aim was to interview three participants in each case study school: the science specialist, a teaching colleague, and head teacher or another member of the senior management team as well as observing a selected lesson taught by the science specialist. However, due to staff turn-over, absence on the day of the visit, or other unanticipated event that prevented staff from meeting the evaluators, it was not possible to carry out all three interviews in all of the 25 schools. Table 37 shows the number of people interviewed in each category of school.

Table 37. Number of individuals who were interviewed

Participants	Full CPD Group	Partial CPD Group	Control C
Science Specialists	8	7	7
Teaching Colleague	9	5	8
Senior Manager	7	6	6
TOTAL	24	18	21

The lessons observed were those usually taught by the science specialists. Full details of the participants and the data are provided in Appendix E. We would emphasise at this point that as the focus of this study was to evaluate the impact of the CPD, our focus in the discussion below is on our observations and discussions within the case study schools of the full and partial CPD groups. Where we feel that it is relevant to include information on the control group schools we have done so, but we feel it important to make clear that there was no clear distinction, in terms of what we saw or heard, when we went into schools on the basis of the group they were allocated to. Indeed, one of the most exemplary examples of science leadership by a science specialist was in a control group school (21C) where a very charismatic science specialist was working with all staff to develop the teaching of science. Furthermore they were aware of current research in science education, the evaluation team newsletters were on the staff notice board and staff were familiar with the project and had clearly talked about it, and the science specialist had started up a thriving after school science club that, due to increasing pupil numbers, now had to run on two evenings per week rather than just one. Furthermore, despite feeling disappointed not to have been allocated to either

CPD group, the head teacher was fully committed to remaining in the study to the end and shared the science specialist's vision of science as a key subject within the school.

The data are analysed and discussed under six headings:

- The science specialists' experience of the CPD
- The impact of the CPD on the science specialists
- The ways in which the CPD was shared within and beyond schools
- The impact of the CPD on the school's science curriculum
- Other factors of influence
- Participants' views on the experience of being part of an RCT

A note on the coding reference scheme:

The following coding reference scheme has been used. Schools are identified using the number assigned to them for the randomised controlled trial followed by the letter denoting group membership:

- Full CPD group is A
- Partial CPD group is B
- Control group is C

Hence, *43B* refers to school 43, which belonged to the partial CPD group. Participants and excerpts from their interviews are tagged by an alphanumeric identifier with the following format: participant category/school identifier/group membership. For example, SS*4A* refers to the science specialist (SS)from school 4 in the full CPD group, *TC78B* refers to the teaching colleague (TC) from school 78 in the partial CPD group, and *SM22C* refers to the senior manager (SM) from school 22 in the control group. Reference to excerpts in the field notes follows the same alphanumeric format, except for the first two letters; that is, FN takes the place of the letters denoting the participant. FN78B, for example, refers to the researcher's field notes from school 78 in Group B.

11.2. Science specialists' experience of the CPD

11.2.1. Views of the CPD course

All eight science specialists in the full CPD group schools expressed their feelings of enjoyment and satisfaction about the CPD course:

> *I really enjoyed the process of the course … It definitely exceeded my expectations, which was great. SS4A*

> *It was good quality CPD without a doubt. SS34A*

The science specialists gave two main reasons for this positive impression.

Firstly, the course provided them with useful resources and ideas, including directing them to teaching materials of which they were previously unaware:

> *We were given lots of tips on where to find things like ... our last session was on science through stories and we were given the names of books ... really good materials that you wouldn't necessarily know about if you hadn't been told about them. SS58F*

> *I got lots of ideas about how to make [science lessons] more interactive … every now and then I find myself remembering things from the training and using them. … I hope that's had an impact on my practice and it is more hands-on than it was maybe. SS34A*

Secondly, the science specialists encountered ideas that aided them in their role as subject leaders, for example, acting as coaches to other teachers in their school:

> *There has been an emphasis on coaching, how to support colleagues in a one-to-one situation and to empower them to be confident in Science and the subject knowledge, so, yes, and able to apply it and help my colleagues here. SS4A*

The science specialists also took the opportunity to build networks during the residential periods, twilight sessions, and regional and local meetings. The course facilitated the development of more expansive networks, which served as post-CPD support:

> *… now there's networks that I've built … within science that are out of this region… that I can email and go, "Oh, do you remember that on that course? Do you have a copy of it?" or, "Have you got this?" or, "Can you remember what we did?" … so I've that wider network now as well. Which … is a great resource to have. SS86A*

One science specialist focused on some significant elements of the course that other participants had not mentioned. It is apparent, from the excerpt below, that this distinct learning experience had an impact on this science specialist's practice:

> *It was nice to be out of school ... going somewhere … to learn all these experiments and have a go ourselves in an environment that was comfortable to us … and for us to do it with other adults and basically all we were doing is what I now do with my children, … in terms of "Ooh, well let's just have a go, have a go, try," and [the course facilitators] were doing that with us really, … and then asking us the questions, "Well, why do you think that's happened?" And now I do that with my children in my class. SS86A*

Full CPD group science specialists found two aspects of the CPD course problematic. Four of the eight science specialists commented that, during the first residential period, the science content that they were required to learn was of a higher level than they had expected. For example:

> *Subject knowledge was, some of it was up there … because that wasn't just me, that was lots of people saying that. It was more KS4 perhaps than KS3… SS94A*

Two of the eight full CPD group science specialists expressed the view that the coursework during the residential period was excessive:

I found the residential periods quite intense … it felt like we were cramming and … obviously, there was a lot to get through... SS34A

As a result, they felt pressured, especially when they went back to their schools, where they had to do the coursework on top of their teaching responsibilities.

All partial CPD group science specialists who attended the CPD course had positive views about the CPD course:

The science specialists valued the focus of the Group B CPD on 'how to teach':

…more teaching ideas … it's helped because it's given me more ideas… SS78B

Just lots of, like, little ideas that they show us on the course just to get children to start thinking in a scientific way, … actually giving them like things to stop and think about, I think that was the main thing for me. SS78B

There was a mix of views on the comparatively light content on subject knowledge in the partial CPD group. On balance, the science specialists felt happy with their level of CPD despite the lack of teaching on subject knowledge, which some felt they could study on their own:

… I think that the knowledge side of it which seemed to be – that's what seemed to be what the other group was offered I could find out for myself I felt [This teacher believed that the main difference between the full and partial CPD groups was that the full CPD group received more CPD relating to subject knowledge which, had they wanted, they could have found for themselves]. *So I was more interested in the practical side of it I think and I wouldn't have wanted to spend that long outside my class really. SS85B*

11.3. The impact of the CPD on the science specialists

The science specialist, a teaching colleague, and a senior manager of schools in both CPD groups were asked about their perception of the impact of the CPD on the science specialist.

Responses clustered into three areas: personal benefits (e.g. confidence in teaching science, improved subject knowledge), benefits to teaching (e.g. changes in classroom practice), and subject leadership.

In general, the full CPD received more mentions of impact than the partial CPD. Of the various aspects of impact, two were consistently cited by more than half of the science specialists in each intervention group: confidence in teaching science, and competence and confidence as subject leaders.

11.3.1. Confidence in teaching science

The confidence that the science specialists reported developing as a result of undertaking the CPD courses was apparent in different ways. Some science specialists derived confidence from having the feeling that *"we were a step ahead" (SS75B)* and from learning a repertoire of teaching strategies from the CPD course *(SS97B)*. Another gained confidence by comparing her practice with the ideas about good practice from the CPD course:

> *... it sort of gives you an idea as to what is good practice and if you are actually doing that sort of thing anyway, it gives you confidence doesn't it, because you know you're working in the right way, yeah? SS85B*

This particular science specialist also had opportunity to disseminate ideas from the CPD course to a local network of schools. She compared herself to teachers from other schools in her area and found that:

> *I'm much more confident in delivering science than they are and I think that's to do with the CPD. SS85B*

Other science specialists gave specific evidence for their confidence. One felt *"more confident in… using the science equipment correctly and accurately" (SS4A)*, while others felt that their increased confidence had translated into changes in their teaching practice, often in implementing open-ended learning activities in class:

> *I'm more confident definitely … and I'm happier to take a few more risks as well so I mean I think like [the open investigation] this afternoon could be a bit of a risk in some ways but … I remember doing it last year and it was chaotic. SS34A*

> *I think in the past I'd teach science but I'd just teach them what it was I wanted them to know, whereas there it was very much getting them to be much more open-ended with things. … And like with the experiment we've just done now, that kind of thing where they're not really getting an answer, so now they've gone away and they still don't really know ... And it's like having the confidence more to do that, and I think that's what I've benefited from. SS78B*

11.3.2. Competence and confidence as subject leaders

Science specialists reported that the CPD had increased their knowledge and skills in how to perform their role as subject leaders. One new subject leader, who succeeded the science specialist who attended the CPD course, commented about the science specialist:

> *… [The CPD course] improved her knowledge and understanding and I think because of that she was able then to support other staff in school better… SS79A*

One science specialist found the discussion about specific aspects of being a subject leader from the CPD course helpful:

Well it sort of – we talked about how to do a book scrutiny, how to do a lesson observation, how to plan Science, all those sorts of things and you know … so yes I would say so. SS85B

For some science specialists, the knowledge and competence translated to confidence:

… I feel a lot more confident in advising other teachers about science … SS86A

… we have a science specialist who is on the ball now and knows what she's talking about and has had training which we've never been able to do before and in terms of bringing the new curriculum in, I think that's supported [her] in feeling confident. SM78B

For other science specialists, this confidence was evidenced in their assertiveness in their role as subject leaders. In one school, for example, the improved competence of the science specialist as a subject leader (see first excerpt below) had been interpreted by the teaching colleague as a marked change in the science specialist's leadership style (see second excerpt below):

I do look at planning with other teachers and so I would have input there in what I think they should be doing, so give them ideas for what they should be doing. SS58A

I think she is more aware of what she is doing, she's got more direction … whereas before it was more guidance. Now it's help, as well as, 'this is what you're doing.' And if I go to her now looking for help she can tell me, 'You can do this, you can do that, you can do this type of thing.' Whereas before it was, it took more of a discussion rather than a leadership role. TC58A

Other science specialists had the confidence to push for the implementation of ideas. One science specialist described that the CPD course gave her a more authoritative voice in promoting a particular monitoring practice:

… I suppose, me coming back and saying: 'I think it would be a really good idea to move forward with this' … because I think I could come back and quite legitimately say: 'I think we do need to continue working on this. Let's try it on Sc1, and then roll it out and see how it goes. SS3A

Generally, there were very few teaching colleagues who commented on the impact of the CPD on the science specialists. The few comments were confined to the competence and confidence of the science specialist as subject leader. For instance, according to one teaching colleague, one facet of the subject leader's role that the CPD course had affected was the science specialist's competence to lead with respect to the implementation of the new curriculum:

… it's been definitely valuable from a whole school perspective I would say. … in her roles as science specialist, [she is] more familiar in how things like the new curriculum might be changing, of things that possibly she may not have been quite a

confident herself. When she's come back you can see that as the science specialist she's a bit more up-to-date on the changes. TC97B

The CPD appeared to have influenced the science specialists in other ways. These are discussed below.

11.3.3. Impact on subject knowledge

CPD from the longer course but not that of the shorter partial CPD, had a perceived impact on the science specialists' subject knowledge. One science specialist in the full CPD group made the specific point that:

> *… it's brought things that I've learnt – I mean, I did my O-levels in the 70s, and it was very much physics, chemistry, biology, we choose one and that's all we ever did. It's been quite nice to refresh when they've been asking questions about various things. Sort of, lurking, I've been able to pull that … and then it's, sort of, built on that knowledge, and thinking things that we weren't quite sure … so it's – yes, it's answered a lot of misconceptions that perhaps we had at grammar school. SS3A*

She did, however, express doubt over whether the improvement in her subject knowledge would be reflected in the test:

> *"There's lots of things that I feel I now have stored in here [taps head]. Whether I'd be able to get what I know down on paper." SS3A*

11.3.4. Impact on teaching approach

The science specialists reported a change in teaching approach that featured more practical, hands-on, open-ended, outdoor, and inquiry-based science activities.

One science specialist used ideas from the full CPD course to *"make [science] more interactive"* and reported that his practice was *"more hands-on than it was" (SS34A).* Another science specialist described what seemed to be a strong impact of the partial CPD programme on her use of the practical approach: *"It's forced me to be more practical and to feel guilty if I'm not practical" (SS78B).*

Other science specialists indicated that they had previously had investigations and practical lessons in science, but the CPD influenced their delivery of these lessons in various ways:

> *[The pupils] were always able to investigate a lot, they do spend a lot of their time doing that. I suppose my questioning has changed, my way of eliciting answers from children and pushing them on to give me what I'm looking for and what I know they have, I think my questioning has improved, and ways of finding out what children know and what they need to know next and how I can give that to them. SS58A*

Field notes from lesson observations made on full CPD group science specialists' classes offered some corroborating evidence. The science lessons that were delivered by the science specialists who reported a change in practice featured practical experiments *(FN86A, FN58A)*, various aspects of scientific inquiry such as observations and prediction *(FN78B)*, hands-on explorations *(FN34A, FN97B, FN3A)*, and open-ended activity *(FN94A)*. Furthermore, the focus group interviews with their pupils elicited reports about science classes where they performed experiments *(FN34A, FN86A, FN78B, FN97B)*, practical work *(FN3A)*, and hands-on activities *(FN94A)*. One group of pupils, though, talked of science as consisting not only of experiments, but also of thinking, talking, and writing *(FN78B)*. They also said that *"science is about 'discussion and finding things out" (FN78B)*. It should be noted at this point that teachers in some control schools also made use of open-ended activities (FN21C) and hands-on explorations *(FN21C, FN72C, FN74C)*. Furthermore pupils in control school focus group interviews also reported doing lots of experiments *(FN21C, FN38, FN72, FN87)* that they enjoyed.

Two groups of pupils compared their science classes during the current and previous academic years. A group of pupils said that they were doing *"more practical work"* this year with the science specialist than in the previous year, although they did suggest that this might be because they were older *(FN3A)*. Another group of pupils also reported that they were having *"more experiments … than last year" (FN86A).*

In contrast to the change described above, teacher-participants from two schools in the partial CPD group reported considerably less change. One science specialist "*did not feel that her science teaching had changed in any particular way" (interviewer's notes)*, but then described a change that was driven not by the CPD course but by the school's move towards *"much more of a cross-curricular direction" (SS75B)*:

> *The previous system was inquiry-based, and this is still a key part of lessons, but we have probably introduced more facts / knowledge. SS75B*

In another school, the science specialist and the senior manager both said that, for a time, there was a surge in the use of the practical approach and investigations in science teaching, which was an impact of the partial CPD course. However, this impact became less noticeable when the school changed its approach in an attempt to balance the focus on 'inquiry-led science' with a focus on 'knowledge-based science':

> *The [PSS] CPD did impact on [the science specialist's] teaching. In general, we introduced more skills-based teaching, practical, inquiry-led science. We already had this, but there was more in our science teaching. Now, however, we have moved back slightly to knowledge-based science. … Don't get me wrong, we still have practicals and we still do experiments, but knowledge is emphasised too. SM52B*

The above case underlines the fact that the long-term impact of the CPD course will be influenced by various factors at the school and individual levels.

11.3.5. Impact on the science specialists' assessment practices

One science specialist commented specifically on the impact of the CPD on their assessment practices, reporting improvement in their competence and confidence to perform their monitoring and assessment function as subject leaders:

> *… [The CPD course] has contributed to improving my confidence. I continue to monitor colleagues' and pupils' work. I discuss with colleagues how they can address the Learning Objectives. I am probably doing more analysis of pupil results than I was last year. SS52B*

11.4. Sharing of the CPD within and beyond participating schools

One aim of the CPD was that participating teachers should cascade their learning to other teachers in their school, and to other schools. This was achieved through a variety of means, including: staff meetings/sessions, in-school training sessions, informal talks or conversations, coaching (one-on-one or small group), and local area cluster group meetings. We would add that even without undertaking this specific CPD some of the science specialists in control schools supported their colleagues by sharing ideas and practice and the level of support offered by science specialists seemed to depend more on them as individuals than on the group they were in. A keen, highly motivated and charismatic science specialist, such as SS21C was doing more to lead science in their school than a science specialist in the full CPD group who had little intrinsic motivation towards science and who had been sent on the CPD by their head teacher with little discussion.

Results were variable. Most schools reported benefits from sharing information, though many also identified problems and barriers. For example, one science specialist's response suggests that the process of sharing the CPD required time to reflect on what elements to share and how to share them, so that other teachers would recognise the applicability of these ideas in their own particular settings. Finding the time for reflection was seen as an added burden in view of competing priorities, as the science specialist described:

> *… I've got all this stack of information and I need to do something to it before it's in a state to pass on. … I want to be able to do it in the most effective and the quickest way and everything because … you can't literally just go, "Well here you are then"…. Because it's of no use to anybody like that. And I need time to be able to, you know, work out how to do it and what's the best way to do it, and of course that type of thing never gets very high up on the list of priorities. SS94A*

Even with planned dissemination, the response at the school level could be difficult to predict as this science specialist's description of her teaching colleagues' response shows:

> *But it is a frustration, … I did a session with all the staff on investigations, … I was very keen on the one that we used on the course which is all the 'Post-its' that you move down and all that, I found that really helpful, so I'd used that in class last year*

and … the children seemed to really like it … but the staff, there was just this real reluctance to use it. … And there was like two members of staff who did eventually start, once they'd started, the others actually did, but I felt that there was a, "I don't want to do something wrong, I don't want to," do you know what I mean? … once we got started, …. [the] reception [teachers] were saying, "That'd be really good, the snow thing, to have ... I think it is a matter of confidence, … and there was also I think a reluctance possibly to change. SS94A

One school reporting little impact had relied on word-of-mouth dissemination, assuming that the enthusiasm of the science specialist would lead to more general change. However, this did not happen. In the words of the teaching colleague in the school:

He always came back very animated from them but, no, I can't say I did [see an effect on his teaching] because I wouldn't have been in a position to observe him or discuss it particularly with him either. TC68A

There appear to be a number factors that could affect the dissemination of CPD ideas and their impact at the school level, such as the personal qualities of the science specialist, the nature of their association with their teaching colleagues and, most importantly, the support offered by the head teacher.

11.4.1. What science specialists shared from the CPD courses

Some of the science specialists, the teaching colleagues, and the senior managers in the participating schools described which aspects were shared with colleagues. These fell into four groups: what to teach (e.g. in response to a curriculum change), how to teach (e.g. practical work, investigations, 'wow' starters, working in groups, ways of recording investigations), assessment techniques, and comments on specific resources encountered as part of the CPD.

Two science specialists (*SS4A* and *SS34A*) implemented a whole school investigation using an activity that featured in the full CPD course, the egg drop. This activity seemed to be versatile enough to implement across grade levels, as one science specialist described:

…we did a whole school investigation where we had to – it was dropping an egg from a height ... which was something I just took straight from the course and that was what was interesting. So everyone … from nursery and reception up until Year 6, they all did the same and how they approached it was up to teachers. SS34A

The prominence of the science investigation and hands-on activities in the shared CPD in the two CPD groups seem to indicate that these aspects of the CPD course had a strong and pervasive impact on the science specialists and their schools. The next sections discuss further the impact of the CPD course on the teaching colleagues, the pupils, and schools.

11.4.2. The impact of the shared CPD on the teaching colleagues

Six of the nine teaching colleagues in the full CPD group and four of the six teaching colleagues in the partial CPD group mentioned that the CPD that was shared by the science specialists had an impact on them in terms of either personal gains (e.g. confidence in science teaching, enthusiasm for science teaching, understanding different ways of teaching topics) or influence on their classroom practice (how they taught and assessed science).

Three teaching colleagues in the full CPD group reported changes in their practice. All these changes were linked to science investigations, or inquiry and hands-on approaches in teaching science. Two of them *(TC3A and TC48A)* also mentioned that their confidence in teaching science and their understanding of science as a subject were also influenced by the CPD that was shared by the science specialists. The teaching colleague in one school reported an increased confidence, and said that *"the way to teach science has changed in my mind"*:

> *I feel one of the things that I've learnt the most from … the science leader … is more so not subject knowledge in the sense that I will get a higher score in the test, but actually how to conduct a science experiment. So the importance on scientific inquiry, predictions, evaluating and that constant 'why'. Why is this happening? … My understanding of science now, is that, it's for the children to explore and for children to be very hands-on and practical. TC48A*

The teaching colleague in another school, a non-science specialist, learned from the science specialist the importance of giving attention to the *"different sides of science"*, namely, *"physics, biology and chemistry"*, rather than just focusing on biology, which was her preference *(TC3A)*. She also told about how the coaching that she received from the science specialist built up her confidence in teaching science:

> *… but actually coming in here and being able to look at how [the science specialist] plans science and I've had a go at it on my own, and her giving feedback to me and saying, "That bit's really good, why don't you try this next time." TC3A*

The teaching colleagues mentioned three elements that contribute to achieving impact from the shared CPD: observing the science specialist's practice, trying out the CPD in their own practice, and getting feedback and direction from the science specialist.

Three teaching colleagues in the full CPD group adopted ideas and resources on assessment delivered in the course. Two teaching colleagues *(TC58A, TC86A)* adopted a CPD resource of tables that were helpful in assessing pupil attainment levels. Another teaching colleague described how, by referencing an idea from the CPD course that the science specialist shared, she obtained support for her move to use alternative recording strategies (e.g. photographs, videos), rather than the head teacher's prescribed method of producing written records of investigations.

In another partial CPD school, the change in the colleague's way of teaching science was in support of the science specialist's move for a more practical approach and was for the purpose of maintaining consistency in how the pupils were taught:

> *… [The science specialist] wants science teaching in a particular way in school. … So I've got to sort of toe her line in that respect …. So, yeah, it has changed the way I teach science. It's more practical than it ever used to be … Once [the science specialist] started doing it, I made a concerted effort to say, "Right, you know, I've got to make time … and children are more engaged and it's more interesting ... And it's more fun. SM78B*

11.4.3. Impact on pupils

The participants were asked about the impact the CPD courses had on pupils' attitudes, motivations, and learning outcomes.

Almost all the impact on pupils described by participants in both CPD groups related to hands-on, practical, open-ended, investigative, and outdoor learning activities. This is likely to be linked to changes in teaching practice focusing on practical, hands-on and inquiry approaches. For example, in one full CPD school, the science specialist introduced the teaching approach that he learned in the CPD course in his science class:

> *… basically all we were doing [in the CPD course] is what I now do with my children, you know, in terms of "Ooh, well let's just have a go, have a go, try," you know, and they were doing that with us really, you know, and then asking us the questions, "Well, why do you think that's happened?" And now I do that with my children in my class." SS86A*

According to the teaching colleague, the science specialist pushed for the implementation of two ideas from the CPD course in the science classes - the 'wow' starters and practical approach:

> *… starting our lessons with a wow at the beginning, that's been really influential to engaging the children … and that's the main thing that's been embedded into our learning now which has been brilliant because it engages children at the start of every topic that we do. TC86A*

As a result of these actions, the science specialist reported that the impact on the pupils and the entire school:

> *… because of the course that I've been on and I kept coming back with all these ideas, I would say that's kind of reflected throughout the school, and there is a real buzz about the subject. I feel the children enjoy science again. SS86A*

He and the teaching colleague then pointed out two pieces of evidence to support this — *"improved quality from [pupils'] work in books" (FN86A)* and the high rate of participation in an after-school science club:

I think, in terms of the quality of the work in the books as well really, you see that now, you see the enjoyment with the experiments, because … there's not as much pressure as to this is right, this is wrong, because we've gone done the inquiry route and the children justifying their own enquiries at their level. SS86A

…the club in terms of the intake we've never had such a high intake… for the after-school club … in particular again because it was after school.... TC86A

Although at this point we return to the fact that SS21C had also set up an after school science club that was very popular and had to be expanded to two evenings per week in order to cope with growing demand which suggests that it is the personal drive of the science specialist that makes a substantial contribution to changes within a particular school.

The teaching colleague predicted that the impact on pupils' learning outcomes would come through the thinking and questioning skills that pupils develop as a result of engaging in science investigations:

[On whether the shared CPD made a difference in pupils' learning outcomes] Yeah I think [the pupils are] engaging because they are asking the question that's getting them to think about well why is that happening? What is that? It's getting them to question and if they get them to question and get them interested, then, you know, I mean it follows they'll be learning about it. Because it's exciting and engaging them, wanting to ask the question, they want to know why, so then I think it follows that it will, that the learning will come through. TC86A

The lesson observation with the science specialist's class and the focus group interview with the teaching colleague's pupils provided some corroborating evidence of pupil engagement, though learning outcomes could not be judged since the lesson did not reach its conclusion that day.

11.5. The impact of the CPD on school science provision

The majority of the participants in both CPD groups reported that the importance of science had been maintained despite the fact that science lessons in primary schools have been given a lower status since national curriculum tests for 11-year-olds in the subject were scrapped. Indeed, some head teachers felt that whilst they strove to ensure that science was timetabled every week and, where possible had an increased time allocation. However, it was felt that this was difficult to justify to governors given that Ofsted inspectors appeared to see English and maths as the sole key subjects.

Participants were asked about plans or actions that were currently in place for science, and how science featured in the school action plan. The participants, in all the groups, mentioned actions or plans that pertained to teaching, science in school, teacher support, and pupils. There was a variety of actions and plans, with those most frequently mentioned being implementation of the forthcoming new national curriculum, establishing cross-curricular links between science and other subjects, and initiatives such as ‘science weeks’. In particular, the CPD was seen as very helpful in preparing for implementation of the new curriculum:

> *Because of this project we’re much further ahead for the new curriculum 2014, we’ve already changed our schemes of work and so on, in order to teach to the new requirements … SM85B*

Participants were asked about the management support for science in their schools. In general, science specialists felt they were well-supported, mentioning most frequently support for participation in CPD, support for the science specialist, and resources and budget allocation.

One question that was posed to the senior managers was about the nature of the role of the science specialist in their school. The senior managers elaborated on the following aspects of the science specialists’ role: lead the subject, support teaching and learning, and possess certain knowledge and attitude. The specific aspects of the role that were most mentioned were responsibility for maintaining a high profile for science, responsibility for implementing new teaching approaches, and monitoring and evaluating the various elements of science teaching and learning. Participation in CPD was seen as a key element of the role.

The science specialists and teaching colleagues were asked about the science methods they used in science lessons, while the senior managers were asked about how science is taught in their schools.

Virtually all participants reported using methods that belong to a wide spectrum of inquiry-based teaching strategies such as teacher demonstration, online virtual experiments, exploration, discovery learning, teacher-guided investigations, fair testing, and open-ended investigations. Almost all participants reported using methods that feature less teacher control and more pupil participation such as practical work, hands-on activities, partner talk, as well as inquiry-based methods.

At the teacher level, six teacher participants reported a change in their teaching methods. However, some of these were not a change of methods but were actually a change in focus or approach in using a particular teaching method as the following evidence suggests. One teaching colleague reported a change in the way she approached the teaching of investigations but could not say *‘definitively’* if that change had been due to the CPD shared by the science specialist:

> *I’m not sure if it’s part of the CPD. But even over the last year my change from just doing investigations and what have you to doing investigations and how can we change, like looking at the investigation more and analysing it and how we can change the results by changing a certain aspect of an investigation. TC58A*

Another, however, changed from the *‘safer*’ practice of whole class demonstration and using fictitious experimental results to open-ended practical work, a move that was encouraged by the science specialist who attended the CPD course *(SS79A)*. One science specialist explicitly attributed the change in his practice to the CPD course:

... it is more hands-on than it was maybe ... I think before everything had to be a fair test and that was you know ... that was one of the things I got from [the CPD] – you know sometimes it's just observing. SS34A

At the school level, the CPD courses also brought about changes in science teaching in the schools. One change that was reported was in the quality of lessons. Two science specialists reported that *"the way that we are teaching it is more exciting" (SS58A)* and that adopting pupil-centred discovery made *"the quality of the lessons ... more exciting for children" (SS79A)*. In another school, the school-wide adoption of "*starting lessons with a 'wow' at the beginning"* and doing *"the practical part of science [and] ... the experimental part"* in science lessons were some of the school-level changes in science teaching that one teaching colleague attributed to the science CPD shared by the science specialist *(TC86A)*. The senior manager from this school offered another perspective to the impact of the CPD on the science teaching in their school:

[The science specialist has] helped us to focus more onto the scientific element of that .... Yeah, subject knowledge .... Because not everybody has science as a specialism ... I think [he has] helped us realise that anybody can teach it as long as you've got this, this and this ... and you engage and enthuse. SM86A

11.6. The influence of other factors, including other science CPD

The science specialists and teaching colleagues were asked about other science CPD that they undertook within the last year or so outside of the project. The number of science specialists and teaching colleagues who had undertaken CPD courses on various topics are summarised in Table 38.

The most frequent focus of other science CPD courses undertaken by the science specialists and teaching colleagues was the new science curriculum, followed by science investigations/practical science.

Table 38. Other science CPD attended by the science specialist and teaching colleagues

Focus of CPD	Full CPD Group		Partial CPD Group		Control Group	
	SS	TC	SS	TC	SS	TC
New science curriculum	5	0	1	1	1	0
Science investigations/practical science	0	2	0	1	0	4
Other or unspecified topics	1	1	0	1	2	3
No other science CPD	3	6	4	2	2	2
Number of Participants asked	8	9	5	5	6	7

Knowledge of the topics of the other science CPD was pertinent to the issue of whether the other CPD could have confounded the effects of the PSS CPD, both in the tests of subject knowledge and in other areas of possible impact. None of the 'other or unspecified topics' in the table above appeared to focus on subject knowledge, so it is unlikely that the outcome of the RCT was affected. Other CPD courses may, though, have been partially responsible for some aspects of impact that were attributed to the PSS CPD. For example, the courses on science investigations and practical science that were undertaken by some teaching colleagues could have contributed to their greater emphasis on inquiry and practical approaches in teaching science, which they reported as impact of the PSS CPD courses that they had undertaken. Furthermore, two science specialists in the partial CPD group attended courses delivered by people who had either worked closely with the Wellcome Trust or the same people who delivered the PSS CPD course. The interviewer's notes on science specialist SS*52A* provided this explanation:

> *[The science specialist] felt that the [PSS CPD course] training (which took place more than a year ago) was very useful. But the school had also had several days input from a primary school science consultant (also more than a year ago). It was evident that this consultant's work had had more of an impact than the [PSS CPD course] training. However, the consultant was [reported to be] working closely with the Wellcome Trust, so there was some overlap between the two sources of CPD, and this was difficult to disentangle.*

Another science specialist talked about the overlap between the PSS CPD and the other CPD in which they had participated, and their explanation of the impact of the PSS CPD and these other courses further highlighted the difficulty in singling out factors that produced impact:

> *Interviewer: So changes to the curriculum which you've brought into the school and I've gathered from speaking to people you've been doing that very effectively, they didn't come from this CPD, they came from the ...?*
>
> *SS97B: No, I'd say both.*
>
> *Interviewer: Both?*
>
> *SS97B:I think mainly it was from this [PSS], because that was the one they kind of highlighted first, and then I've been on the other ones since those or in between them, so ... This one has kind of built each time we've been on the course, like each day we've been on it's kind of built as they know more about it and they've done more work on it.*

The aspects of impact of the other science CPD that were reported by the participants in the control group were no different from those that were attributed to the PSS CPD by the participants in the two CPD groups. However, there was no evidence that being in the control group motivated the teachers to seek actively an equivalent CPD experience.

In their interview, the participants mentioned several factors that they felt could have influenced the dissemination and impact of the CPD courses, or that could be considered as contributing factors for some of the reported or observed impact.

Teacher turnover was seen as a factor that could adversely affect the impact of CPD although this was not quantified. In one school *(48A)*, for example, the science specialist had already left the school by the time of the interview. A new science specialist had taken over, but according to the teaching colleague, the new science specialist had her own vision for science. Without the transfer of vision from the old to the new science specialist, the gains due to the CPD could be lost:

> *I think for us as a school, it was a shame that [the CPD science specialist] left when she did ... Because I think the impact and, as you've used the term, legacy of this, the legacy would have lived on for a lot longer, and science would have gone from strength to strength ... while [the CPD science specialist] was here, I did feel part of this, because she was always talking to me about what's happening and giving me ideas, whereas that has taken probably a dip, because [the new science specialist has] not been about. TC48A*

Personal preferences and attributes were also seen as potentially affecting the dissemination of ideas and resources from the CPD courses in the participating schools. In one partial CPD school *(80B)*, for instance, the senior manager reportedly had difficulty trying *"to get the [science specialist] to disseminate [the] CPD into [the] school"* because the science specialist *"seemed to see [the CPD] as personal professional development"* rather than something to be shared in school *(FN80B)*.

In contrast, another senior manager *(SM86A)* commented that the science specialist was the type of person who *"goes the extra mile"* even in the way that he shared the CPD in school. The *"general staff opinion [was] that [the] science profile has increased as a result of the CPD" (FN86A)*, and the senior manager expressed satisfaction with the outcome:

> *I don't think at first as well, I don't think we realised how much more we'd get from it as a school and as a staff. I think we appreciated it, it was part of [the science specialist's] mission and CPD, but I think we have been very impressed by, you know, what he's gleaned from that and been able to pass on. SM86A*

Some participants mentioned factors that could explain, partly or wholly, the changes to their practice or personal qualities. For instance, for some teachers, the more they were teaching science, the more confident they became. Hence, teaching experience was the reason for their improved confidence in teaching science.

School-level factors appeared to influence some of the changes that took place in the school during the duration of the CPD project. These included changes in school policy on how science was taught (e.g. through a more cross-curricular approach), changes in monitoring and assessment practice, and the introduction of science-focused school targets aimed at improving pupil outcomes.

11.7. Participants' experience of the evaluation

This section considers participants' views of the nature of the data gathered in the evaluation, and their views more widely on the experience of participating in a randomised controlled trial.

11.7.1. Participants' views of the data gathered in the evaluation

Participants were asked for their opinion on the suitability (face validity) of the subject knowledge tests and the teacher confidence and pupil attitudes questionnaires to the teachers and pupils as part of the project evaluation.

The majority of the participants felt unable to comment, did not give an answer, or gave a qualified response. The few participants who had reservations about the tests and questionnaires gave a number of reasons. These included teachers feeling that the test of subject knowledge was "more of a test on my retention of things since high school" (TC79A), and that pupils' attitudes may have been more appropriately gathered through talking to pupils, as they can talk more easily than write [we would note here that we did in fact also gather such views from the pupil focus groups in the case study schools].

The point that received the most comment related to subject knowledge tests with twelve teaching colleagues and five science specialists commenting that these were difficult and/or "pitched higher than the standard that we teach in primaries" (SC18C). One science specialist, who attended the full CPD course and presumably had subject knowledge training from the course, described his experience with the tests:

> *… when the first paper came out it was pitched at more at secondary school … there were questions in there I couldn't do, … that I found hard … and I don't know if it was pitched even higher than secondary school ... the viewpoint from everybody on the course, quite a few people comment about the exam papers and couldn't see the connection to how to, how does that measure what that course has done for us. SC86F*

Some teachers felt that the tests measured a narrow aspect of teacher knowledge, and failed to reflect the contingent knowledge that teachers gained in practice.

In contrast, one teaching colleague *(TC58A)* reported that taking the tests inspired her to improve her subject knowledge, such that whatever improvement that might be reflected in the second tests could be attributed to the studying she did.

There were mixed views on the coverage and level of science content in the pupil tests, with some feeling they were fair, others saying that the tests were better suited to *"higher science level children" (SS46C)* and others suggesting the tests were *"a little bit more lower and easier for them and obviously directed at them" (SS31C).* Some teachers also reported about the discrepancy between the science content in the tests and the science content that they had covered in class.

Other teachers observed pupils who did not *"read the question properly" (SS46C)* and pupils who did not perform well under time pressure *(SM78B)*. Some teachers also commented that their pupils would not be familiar with taking such tests. Advocating other means of gather data on learning:

> *So, I'm not sure whether the tests that they've done will reflect their knowledge and understanding. Probably their books, if you know, if you looked across books and spoke to children in an informal way, I think you'd have a better understanding of what they knew and didn't know than what a test shows. SM78B*

> *I think the best way to measure the children's learning is to watch them in a science lesson as opposed to see what they – what score they would get on a test. TC48A*

Other factors mentioned by participants that could have affected the pupils' test outcomes were the changing curriculum, and the quality of teaching in the science classroom.

11.7.2. Participants' views of randomised controlled trials in education

Of the sixty-three participants, twenty-seven (9 full CPD group , 7 partial CPD group and 11 control group) reported that the demands of participating in the evaluation RCT were not onerous, and simply subsumed them into their normal duties.

A quarter of the participants made the point that they would positively consider participating again in a RCT, with half of these indicating they would participate even if assigned to a control group. The other participants tended to be indifferent.

Some of the participants who indicated their willingness to participate again qualified their answers. One senior manager said they would make sure that the prospective participants would need to know that there would be a possibility that they would not be gaining anything during the research period if they were assigned to the control group, saying that:

> *I probably would have been less inclined to commit much time and thought had we been in that control project because, as you say, there's not much direct impact on the children in school. SM34A*

A desire to keep up with the participants who would be experiencing gains from the interventions was reported by some senior managers:

> *… we would want to make sure that the quality of our lessons is high because providing a quality experience, an outstanding experience, for our children is what we aim to do. We would no doubt have sought CPD where we felt it was appropriate … SM4A*

Participants who were unwilling to participate again also commented on the perceived implications of being in the control group. One senior manager commented:

> *I would feel that we would be under a certain sense of scrutiny. I would probably be more vigilant around science that I would be …. You'd only undertake [the project] if you were [doing well]. You wouldn't [want to] be [in] the control group to be found out that you were completely inept in science…. I think you would be doubly cautious, and I think you would only say yes if science was on your plan as a development area. SM48A*

The question of how to make participation more attractive for schools in the control group was also posed to the participants. While two participants mentioned *"money to buy resources" (SM72C)* and *"supply money to release the class teacher to do the test … in school time" (TC18C)*, the most common response was a guarantee of CPD at the end of the research period or, at the very least, access to the materials and resources that would be given to full CPD group schools, as was promised in this trial. One participant suggested that information about these expected benefits could convince her to participate:

> *… without actually knowing something, I think that would personally put me off. I'd need to know what was in it for me, what was in it for the school, and more importantly, what was in it for the children ... SS8BP*

Some participants, however, gave the impression that even the promised CPD at the end might not be sufficient incentive to participate. For instance, one senior manager said that if schools had to wait a long time for the CPD school priorities might have changed and the promised CPD might no longer be of benefit to them.

Participants were asked for their views on whether conducting a RCT "to establish the effectiveness of educational interventions is a good way to affect teaching practice", and about the key benefits and limitations of this research approach. Views were mixed, and there was evidence in the replies of only limited understanding of RCTs.

Two senior managers commented positively on RCTs as a basis for making decisions about adopting educational interventions:

> *… randomised controlled tests are the best way of doing things and people don't like doing them in education because nobody wants to be the group that's left out … and it's getting your head round as a school leader the fact that children aren't disadvantaged … So yeah I think it's a good way and then we've got a better evidence base … SM79A*

> *I am aware of the use of RCTs in education and I gather it's quite controversial. But as a scientist I think they are a good idea and we should do more of them. They are robust and they really do tell you what works. I think the Wellcome Trust were right to use this method to evaluate their CPD course. SM43B*

Other responses to the question about the limitations and benefits of RCTs highlighted the lack of knowledge about the methodology, with mention of factors such as failure to take into account teacher and pupil attributes, or the impact of other CPD courses and school-level changes (factors that RCTs are, in reality, designed to tackle).

In the next section we move on to present a summary of the evaluation as a whole.

12. Summary and discussion of key findings

12.1. Introduction

The aims of the evaluation were to provide a good quality evidence base of the school, teacher and pupil-level impacts of the CPD after completion of the course and one year later; and recommendations for the national roll-out of the initiative.

The evaluation employed a randomised controlled trial (RCT) with quantitative and qualitative methods to gather evidence on the impact on science specialists including:

- their science-specific subject knowledge
- their knowledge of how to teach science
- their confidence in teaching science.

To assess the impact on pupils, the evaluation gathered evidence on pupils' science achievement and their attitudes to science.

In order to present a summary of the key findings regarding the impact of the PSS CPD we will, where appropriate, make use Guskey's (2000) five level model in which increasing impact is associated with the higher levels. As can be seen (Table 39) evidence for impact at some levels (2, 5 and to a certain extent 4) were assessed using both quantitative and qualitative approaches whilst impact at levels 1 and 3 was assessed solely on the basis of qualitative data.

Table 39. Guskey's five levels and the methods of data collection

Level	Focus of evaluation	Data collection method
Impact at level 1	Participants' reactions to the course provision	Interviews with SSs, TCs and SMT
Impact at level 2	Participants' learning	Interviews with SSs, TCs, SMT as well as subject knowledge tests and confidence questionnaires
Impact at level 3	Organisational (i.e. school or department) change	Interviews with SSs, TCs and SMT
Impact at level 4	Participants' use of new learning both in the class and in completing tests	Field notes, interviews with SSs, TCs, SMT and subject knowledge tests and confidence questionnaires
Impact at level 5	Student learning and attitudinal outcomes	Field notes, interviews with teachers, pupil focus groups, subject knowledge tests and attitudinal surveys

12.1.1. Level 1: Participants' reactions

Participants were generally positive about the CPD course. Indeed all the science specialists in both CPD groups gave positive feedback about the CPD course. They enjoyed the course and found it useful. The science specialists in both of these groups found beneficial the aspects of the course that dealt with how to teach science and how to be a subject leader. Feedback on the subject knowledge delivered as part the full CPD group CPD course was less positive with many teachers feeling that they were being asked to develop their subject knowledge to a higher level than required for teaching primary school science.

Participants felt more enthusiastic about teaching science and felt they had gained new knowledge and skills as a result of participating in the CPD course. The science specialists' favourable opinion could be traced to their appreciation of the CPD ideas and resources that were related to the teaching of science and their role as subject leaders. Many of the participants were enthused about ideas that were presented on the CPD course on how to teach science investigations, how to engage pupils' attention during science lessons, alternative ways of recording investigations, online resources for teachers and for teaching, how to support colleagues, and how to be a subject leader.

Participants had different views on the value of the subject knowledge component of the CPD. The principal area where there was a perceived mismatch between the participants' perception of their needs and the CPD provision related to subject knowledge. Although it was clear that the science specialists felt that they needed to learn how to teach science and how to be a subject leader for science, they had difficulty making a connection between these needs and the learning of, what they felt to be 'high-level' science content. Some science specialists in the partial CPD group, who had evidently found out about the full CPD course, did not think the missing subject knowledge training a loss because they could find it for themselves if they needed it. These opinions might be an indication that for some primary science teachers what was presented as essential subject knowledge on the CPD course was actually just slightly more knowledge than was needed to teach their pupils. Thus the standard of subject knowledge with which they were expected to engage was far greater than they felt was needed. Indeed there was a view amongst some participants that the PSS course ought to be split between upper and lower primary as the subject knowledge requirements of this two groups were thought to be very different. However such a split, were it to be implemented, might not provide skills needed to lead science throughout the primary school.

The full CPD course had more reported impact than the partial CPD course though the latter was felt by many schools, evidently based on the memorandum of understanding and other information received from the recruiters, to provide the optimum balance between a range of competing factors.

The views expressed by the science specialists raise the issue of the benefits and drawbacks of the longer course versus the shorter course. For example, although the science specialists in the full CPD group spoke of the difficulty in managing the requirements of the course on top of their duties in school and problems with residential periods away from home they also talked about opportunities it afforded them. In particular the opportunity to gain hands-on practice at the NSLC with the teaching methods they would later be applying with their own pupils, the friends they made during the course that became part of their support network, and the inspiration they experienced from interacting with the CPD facilitators. These experiences were missing from the accounts of science specialists in the partial CPD group. However, some science specialists and senior managers in the partial CPD group said that their CPD course was the optimum set-up for their schools. In these schools it was felt that constraints, such as for example the number of staff and availability of funding for teacher cover[19], would have made involvement in the longer course unmanageable. We emphasise here that we make no judgment about the validity of these constraints but merely report the views of staff. This shows that at the school management and individual teacher levels a complex set of considerations would, outside of this pilot study, influence choice-making about the type of CPD to undertake.

For the PSS intervention, the advantages of the longer course appeared to outweigh its disadvantages, especially if assessment is based on participants' self-reporting.

12.1.2. Level 2: Participants' learning

Participants reported gaining new knowledge and skills that improved both their performance as subject leaders, and their classroom practice. The science specialists who attended either of the CPD courses reported multiple layers of learning. First, there were the ideas they learned about how to teach science, the change in their understanding of science as a subject, the improvement in their subject knowledge (for science specialists in the full CPD group), and the knowledge they gained about subject leadership in science. Second, these knowledge gains translated to affective gains such as confidence in teaching science for some and to increased interest in science or science teaching for others. Third, these cognitive and affective gains manifested behaviourally as change or improvement in the way they taught science, in their assessment practice, in their leadership styles, and in the way they supported their colleagues. It should be noted though that, at the level of individual teacher, learning was not as many layered for science specialists in the partial CPD Group as for science specialists in the full CPD group.

Participants from the CPD groups also reported on feeling more confident and prepared for the implementation of a national change in the curriculum. Whilst many of the teachers self-reported an improvement in their subject knowledge this was not supported by the findings from the subject knowledge tests. Indeed, the results showed that despite many participants' beliefs to the contrary there was, in fact, no statistically significant evidence of impact; either in terms of the CPD courses on teachers' subject knowledge as measured by their test results, or in terms of the confidence scores that they gave to their test answers when compared with the control group.

[19] Full supply cover costs were paid for teachers to attend the CPD during the trial. Bursaries are available for maintained schools that cover most costs incurred by staff attending CPD at the National Science Learning Centre.

12.1.3. Level 3: Organisational support and change

Participation in the CPD often resulted in the status of science being raised in schools and increased support for science in schools.

Science enjoyed a high status in almost all participating schools, being seen as the third most important subject after English and Maths. However, we would caution that this might reflect recruitment bias in so far as pro-science schools might have been more inclined to get involved in the project. In many of the schools, the status of science was reported to have improved, and around half of the schools who reported an improved status for science attributed this to the involvement of their science specialists in the CPD course. For example, in two schools in the full CPD group, the time allocated to science lessons had been increased.

The three most commonly mentioned types of support in which senior management teams in the schools extend support to science were through the provision of CPD, support for the science specialist (in the form of support for implementing change as a result of participation in the CPD), and resources and budget allocation. The senior managers, on the other hand, expected the science specialists to keep science in the forefront and monitor and evaluate the various elements of science teaching and learning. Many of the science specialists considered being given time off to do the CPD course as support from the school management. Their involvement in the CPD course, in turn, was cited as the reason for the improvement of the status of science in many full CPD group schools and a few partial CPD group schools. Moreover, changes in school-level monitoring and assessment practices were also attributed to the CPD course in a few full and partial CPD group schools.

The relatively short duration of the evaluation meant that it was not designed to evaluate any evidence about long lasting effects of the CPD and it is therefore difficult to establish possible connections between school-level factors and longer-term impact of the CPD.

12.1.4. Level 4: Participants' use of new knowledge and skills

Participants used a variety of methods to disseminate their learning from the CPD course, with mixed success. The impact appeared to be greatest for the use of practical, hands-on, inquiry-based, and outdoor activities experienced in the CPD. There was very little evidence of dissemination of subject knowledge.

All the science specialists shared the ideas and materials from the CPD course using various means of dissemination in their respective schools, and in some cases, in their local networks. Selected elements of the CPD were shared during staff meetings, in-school training sessions, and during the course of the subject leaders' activities. Virtually all elements that were reportedly shared were ideas and materials for the teaching of science. The most popular element that was shared in full CPD group schools were on investigations, practical and hands-on activities, and 'wow' starters. No-one specifically mentioned sharing aspects of subject knowledge that they might have learnt on the CPD programme.

Many aspects of the impact reported (on both staff and pupils) related to practical, ‘hands-on’, inquiry-based and outdoor activities involving a shift to less teacher control and greater pupil participation. At the school level, the use of ideas and materials from the CPD were reported by science specialists in four of the nine full CPD schools and one of the seven partial CPD schools. There were reports about whole-school activities using ideas from the CPD, the use of the ‘wow’ starters in science classes as well as the adoption of ideas and resources and use of new skills in the teaching of science investigations and for assessment and monitoring. The evidence pointed to much of the adoption of ideas, resources and skills happening during, or immediately after, the CPD.

12.1.5. Level 5: Pupil learning outcomes

The quantitative evaluation of the main study found **no statistically significant** evidence that the CPD produced any impact in terms of measureable improvement in the science understanding of pupils taught by the specialists during the year the course was delivered. However, a **significant** improvement was found in the performance of pupils taught by the science specialists in the full CPD group in the year after the course was delivered. In this smaller analysis of 33 schools, there was a significant difference in the scores of the full CPD group compared to the partial CPD group, and full CPD group compared to both other groups combined (although there were **no statistically significant** differences between the test results for the full CPD and control groups, or for the partial CPD group and the control).

It is difficult to extrapolate why this difference was present, indeed a significance test was carried out for the same schools in the main study and no differences were found thus the difference occurred *only* in relation to test results for pupils in the science specialists’ new classes, and there were no differences in the results for the previous/old classes. One possible explanation for this would be that the effect of the CPD on some of the science specialists had become stronger as they had all completed the CPD and embedded this within their practice by the time they had taken on a new class, one year or so on from completing the full training.

In terms of students’ attitudes the surveys found that there were some instances where the opinions of pupils in the full CPD group appeared to have changed in a different way, or more extensively, than those in the other groups. One particular case related to the statement ‘science has made us better and safer medicines’ in which it was found that the proportions agreeing or agreeing ‘a bit’ with this statement had increased in all three groups, but that the increase in the full CPD group was considerably higher than the increases for the other groups and this difference was found to be **statistically significant.**

Many of those participating in the CPD reported affective gains for pupils, such as improved pupil enjoyment of science, attitudes towards/interest in science, and learning outcomes. Some schools also reported the development of pupils’ thinking skills as a result of the cascade of effects from the CPD although objective evidence was provided to substantiate those claims. There appeared to be an association in participants’ minds between increased engagement and improved learning, although there was no unambiguous evidence either from the lesson observations or the pupil focus groups to support such a view.

12.1.6. Other factors

A number of factors external to the CPD provision appeared to influence its impact. For example some participants reported undertaking other science-related CPD outside of the PSS project that could have confounded the impact of the CPD and also had an impact on their teaching. For instance, other CPD may have contributed to a greater emphasis on inquiry and practical approaches in teaching science. Whilst there was no evidence that participants from the control group explicitly sought an equivalent CPD experience to that of the project CPD, a few participants hinted at making efforts to achieve some gains for science in their schools despite not being in either of the treatment groups.

Overall, the evidence suggests that school-level and individual-level factors, such as changes in staff, whole-school policies, personality and the leadership style of the science specialist, were the factors that were most likely to influence the impact of the CPD. This highlights the complex nature of factors at the school and individual level, making it difficult to isolate aspects of impact that could be unequivocally attributed to the project CPD.

12.1.7. Views on participating in an RCT

Views on the use of RCTs were comparatively neutral, and many participants did not have any idea as to what such a design involved.

Two aspects of the project design (the RCT) seemed to be problematic to the participants. Firstly, there were concerns about the possibility of being assigned to groups whose demands and expectations did not match the school's capacity to commit and its priorities. For instance, there were participants allocated to either the partial or control groups who claimed that they would not have managed the greater time commitment required of those in full CPD group and others, in the full CPD group, who felt that they would have been unable to wait for the full CPD had they been allocated to the control group.

A quarter of the participants indicated a positive willingness to be involved in another RCT in the future (the remainder were indifferent), even if assigned to the control group, and felt that a guarantee of CPD at the end of the research period or, at the very least, access to the materials and resources would have been enough incentive for them to participate.

Secondly, the research period was seen, at least in terms of school planning, to be overly long. While participants recognised the benefit of evaluating the long-term effects of CPD as a better and more effective approach, they also highlight the inevitability of changes in staff and school priorities within that period, which might adversely affect the research design. That said it emerged that many of the participants did not have a clear understanding of the principles that underpin a RCT and, as a consequence, saw the value of their involvement in the project solely in terms of the benefits it offered them and their own school and failed to appreciate the broader implications that might be drawn from the research outcomes and the need for control schools in such studies.

What emerges clearly from this evaluation is the need to consider how the very positive self-reported beliefs made by participants about of the full CPD course can be translated into impacts that can be measured in terms of agreed objective outcome measures. The results of the Extension Study suggest that it might be possible that measurable impact from the CPD, at least in terms of pupil learning, might only be determined when participants introduce new ideas/approaches with a new class. To ascertain whether this effect is reproducible would require a further extension study that focused specifically upon that sub-group of science specialists taking on new classes.

12.1.8. Recommendations for the national roll-out of the initiative

There can be little doubting that the PSS CPD has impacted in a positive way on teachers of science in primary schools. However, if the aim is a national roll-out of the initiative it is important to consider the fact that the intended impact was not consistently reflected in the actual impact in all respects. In terms of ‘soft’ measures the intervention has successfully raised teacher confidence *in teaching science and trying out new ideas*, introduced teachers to resources and materials for teaching (such as those at the National STEM Centre) of which they were previously unaware and, very importantly, helped to develop networks of primary science teachers who will continue to share ideas and resources long after their completion of the CPD. In terms of ‘hard’ measures, such as an improvement in teacher subject knowledge and confidence about that knowledge, the evaluation has found that the CPD (delivered to both treatment groups) had no impact. However, a note of caution needs to be added here as the results from the extension study suggest that such ‘hard’ measures of impact might take longer to manifest themselves. Whilst some statistically significant changes were found in terms of pupils’ attitudes towards some very specific areas of science the value of such change would need to be considered in terms of whether such changes are transitory or enduring over a longer period of time.

On the basis of the results – particularly the lack of evidence of improved teacher subject knowledge (and even with the evidence from the extension study the Effect Size was low) and confidence in that knowledge - our suggestion would be for a shorter hybrid programme. Such a programme, potentially a residential of five days spent at the NSLC, would focus not on attempting to develop teachers’ subject knowledge but on their science pedagogy in terms of specifically increasing their awareness of, and confidence in using, the wealth of science teaching materials that are currently available but about which many of them are unaware.

References

Abrahams, I., & Millar, R. (2008). Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14), 1945–1969.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education:Principles, Policy and Practice*, 5(1), 7–74.

Guskey, T. (2000). *Evaluating professional development*. California: Corwin Press.

Pell, T., & Jarvis, T. (2001). Developing attitude to science scales for use with children of ages from five to eleven years. *International Journal of Science Education*, 33(8), 847–862.

GLOSSARY

CPD Continuing Professional Development. Opportunities for current teachers to undertake course/training to develop their practice.

Group A is the full treatment group (school specialists experiencing the full 24 days of CPD).

Group B is the partial treatment group (school specialist receiving four days of CPD).

Group C is the control group, where school specialists did not experience any of the CPD.

ITT ‘Intention to treat’. The analysis encourages schools to return data even if they have dropped out of the intervention or control condition, and analyses them as if they still belonged to the group they were randomised into. In this study for instance, data from a Group A school that dropped out after one session of CPD will still be included as if they were in Group A.

KS2 Key Stage 2: in England, pupils between the ages of 7 and 11 (Years 3-6).

KS3 Key Stage 3: in England, pupils between the ages of 11 and 14 (Years 7-9).

NSLC National Science Learning Centre.

PCK Pedagogical content knowledge. Teachers’ ability to interpret and transform subject knowledge in a way that facilitates student learning.

RCT Randomised controlled trial. A study in which groups are formed through random assignment to help avoid allocation bias. Typically, one or more groups (the experimental groups) receive the intervention(s) being tested, and one group (the comparison or control group) continues with “business as usual”. Outcomes are measured at specific times and any difference in response between the groups is examined statistically to assess the impact of the intervention(s).

SATs Standard assessment tests. These are tests taken by school pupils at certain times during the educational system as part of the national curriculum. In some cases (e.g. KS3 tests) the test is no longer compulsory in England but previous papers are available for use.

STEM Science, Technology, Engineering and Mathematics.

Appendix A: Defining the expertise of a Primary Science Leader (PSL)

Subject knowledge

A PSL should have a deep understanding of the scientific concepts within the Key Stage 1 and 2 National Curriculum, supported by a working understanding of the scientific concepts within the Key Stage 3 National Curriculum. A PSL should be confident in the use of scientific vocabulary and know how to research science topics and guide their pupils to do the same.

Pedagogical knowledge

A PSL should have knowledge of an appropriate range of teaching methods suitable for the content concerned. Their knowledge should include enquiry-based pedagogies, practical activities, out-of-classroom learning, group work and problem solving, digital technologies, and formative assessment practices. It should also include an understanding of the key features that result in the successful implementation of these pedagogies and how to evaluate the impact of these on their pupils.

Subject leadership

A PSL should have a whole-school vision for science and be able to lead its development by instigating appropriate initiatives, including providing continuing professional development (CPD) to colleagues, monitoring progress and contributing to the strategic development of learning in school.

Appendix B: Outline of the CPD programmes

The full CPD programme focused on subject knowledge, pedagogical content knowledge and subject leadership.

Subject Knowledge
This element of the CPD programme supports teachers to:

- extend and deepen subject knowledge in the *Big Ideas*[20] in science
- understand and manage common misconceptions in science
- model key ideas in science and understand the limitations of common models
- reflect on personal subject knowledge and engage in peer to peer coaching and on line activity
- apply personal subject knowledge to planning activities for school based work
- develop understanding of progression of subject knowledge across the primary age range.

Pedagogical Content Knowledge
Teachers learn about current research and thinking regarding how children learn in science, to extend and enhance teaching and learning across the primary years.

- How children learn and progress in science
- Outstanding teaching of science
- Curriculum design
- Assessment for learning in science
- Using digital technologies to enhance science
- Planning, research and evaluation

Subject Leadership

- Understanding of the role of the Primary Science Specialist
- Challenges facing science subject leaders
- Developing a school development / action plan for science
- Resourcing primary science
- Raising the profile of primary science in school
- Auditing staff needs
- Mentoring and coaching colleagues in school.
- Managing staff
- Planning and leading a science staff meeting
- Work scrutiny
- Lesson observations
- Regional initiatives and networks
- Assessing science
- Special educational needs in science, including gifted and able

[20] W Harlen (ed.). Principles and Big Ideas of Science Education. Hatfield: Association for Science Education; 2010.

Appendix C: Instruments

Research instruments as used throughout the study are included as separate documents. An outline is given here.

Teacher tests

Baseline	BaselineTeacherPaperA.pdf BaselineTeacherPaperB.pdf	These tests were used with science specialists (SS) and their teaching colleagues (TC) at baseline
PostTest1	PostTest1TeacherPaperA.pdf PostTest1TeacherPaperB.pdf	These tests were used with SS and TC at post-test 1
PostTest2	PostTest2TeacherPaperA.pdf PostTest2TeacherPaperB.pdf	These tests were used with SS and TC at post-test 2

Pupil tests

All phases	Year3ScienceKnowledgeTest.pdf Year4ScienceKnowledgeTest.pdf Year5ScienceKnowledgeTest.pdf Year6ScienceKnowledgeTest.pdf	These tests were used with appropriate year groups at each of the three phases (see also Section 5.2)

Pupil attitudes

All phases	PupilAttitudeSurvey.pdf	These surveys were used with all pupils at each of the three phases

Case study instruments

Visit 1	LessonObservation.pdf	This schedule was used in science lesson observation with SS; reused at visit 2
	PupilFocusGroups.pdf	This interview schedule was used with pupil focus groups; reused at visit 2
	SMvisit1.pdf	This interview schedule was used with member of SMT
	SSvisit1.pdf	This interview schedule was used with SS
	TCvisit1.pdf	This interview schedule was used with TC
Visit 2	SMvisit2.pdf	This interview schedule was used with member of SMT
	SSvisit2.pdf	This interview schedule was used with SS
	TCvisit2.pdf	This interview schedule was used with TC

Appendix D: Details of the case study participants and the data

School Identity	Science specialists	Teaching Colleague	Senior Manager	Interviewer's Field Notes
FULL				
3F	✓	✓ Stand-in for TC participant, who was unable to attend the interview	No SM interview (no other details)	✓
4F	✓	✓ (moved from teaching year 2 to year 1)	✓	✓
34F	✓	✓	✓	✓
48F	No SS interview (SS-participant left the school)	✓ (currently assigned as maths SS, but was supervising a trainee teacher who did most of the teaching including science)	✓	✓
58F	✓	✓	No SM interview (no other details)	✓
68F	✓ Not the CPD participant; SS-participant left the school	✓	✓	✓
79F	✓ Not the CPD participant; SS-participant left the school	✓	✓	✓
86F	✓	✓	✓	✓
94F	✓	✓	✓	✓

PARTIAL				
43P	✓	No TC interview (no other details)	✓ (Not the original SM, had involvement with Wellcome Trust and NSLC at national level on the CPD provider side)	None
52P	✓	✓ Stand-in for the TC participant who was unable to attend the interview	✓	Lesson observation only
75P	✓	✓ (short interview)	No SM interview (no other details)	Lesson observation only
78P	✓	(see SM comment)	✓ (also the TC but speaking as SM for the most part; so interview was coded as SM)	✓
80P	✓ Not the CPD participant; the SS-participant left the school and did not share CPD so this SS could not comment about impact	✓ Not the original TC participant; (this substitute TC did not take the tests)	✓	✓
85P	✓	✓	✓	✓
97P	✓	✓ (moved from lower school to upper school)	✓	✓
CONTROL				
18C	✓	✓	✓	✓
22C	✓ (also the SM but interview was coded as SS)	✓	(see SS comment)	✓

31C	✓ Not the CPD participant; SS-participant left the school	✓ Not the original TC participant, who had left the school by this time	✓	None
38C	✓ (attended a different residential science CPD course)	✓ (also a senior manager but interview was coded as TC)	(see TC comment)	✓
46C	✓ Not the CPD participant; SS-participant left the school	✓ (also a science specialist)	✓	✓
72C	✓	No interview (TC participant was unable to attend the interview)	✓	✓
74C	No SS interview (SS-participant left the school; see SM comment)	✓	✓ (SM but also the new SS; took over as SS from SS-participant; interview was coded as SM; did not receive any information about the project from the SS-participant)	✓
87C	No interview transcript (although some notes about the interview were included in the field notes)	✓	✓	✓
89C	✓	✓ Not the original TC participant	No SM interview (no other details)	None

Appendix E: Further explanation of power analysis and modelling outcomes

Details of the power analysis used for the baseline and post-tests are provided in Section 7.2. This appendix provides further detail of what these figures mean. The main details are as follows:

Assuming pre/post correlation = 0.70

Intra-class correlation (ICC) = 0.125

Power = 0.80

Minimum detectable effect size (MDES) = +0.24 / +0.25 between the different treatment groups.

What these figures mean, essentially, is that we have sufficient power in the pupil test results to carry out the required analyses. The **effect size** is the magnitude or impact of a particular variable on an outcome (e.g. the science CPD upon pupils' test results). In our calculations this is expressed as the Minimum Detectable Effect Size (MDES) and it is set at 0.25. This is fairly typical for looking at the impact of an educational intervention. It means that if the CPD has an effect size of less than 0.25 we will not find it - but this is not a problem because it means that its impact is relatively weak and there must be other things apart from the CPD, perhaps in combination, having more of an impact upon the pupils' test results. If the effect size is more than 0.25 we will detect it. In **educational** terms 0.30 would be quite a mild/medium effect, 0.40 would be medium/large, and 0.50 or above would be a large/strong effect. (There is some debate about these effect sizes and these thresholds are not hard and fast. In non-educational studies the thresholds would be higher, often based on Cohen's recommendations for small, medium and large effect sizes).

The **pre/post correlation** is the strength of the relationship between the two sets of tests used - how close and consistent are the two test papers, baseline and post- in terms of measuring pupils' skills, knowledge and ability? The test questions we are using are all from the same Assessment series which the developers had tested and found to have a high reliability (high Cronbach's alphas). Also Years 4 and 5 have done exactly the same test again. So we believe the correlation to be strong and 0.70 is a standard and entirely reasonable figure in this respect. The intra-class correlation (**ICC**) is a descriptive statistic that can be used when quantitative measures are made on units that are organised into groups: in this case the groups are schools or classes of pupils, and 0.125 is fairly typical for this type of calculation.

This information, along with other parameters including the numbers of schools for which we have test papers, is fed into an appropriate software package, and this produces the **power calculation** (the strength of the power calculation is dependent upon the other factors put into the equation). As with the previous round of testing and sampling we have power = 0.80. The power can be expressed as the 'probability of finding a difference', so in this case a difference in pupils' test scores (if such differences exist) between our three treatment groups (full, partial, control). In other words, the MDES and power figures give us guidance on whether we are going to find something out using statistical analyses. The power of 0.8 is high and the MDES of 0.25 is appropriate for looking at an educational intervention. ***So we were confident that we should carry out our analyses because there was a reasonably good chance that if the CPD has had an effect on pupil outcomes, we would find it.*** On this basis we went ahead with our analyses across the three groups (see below) for post test 1 and post test 2, as we did with the first stage of tests.

The question arises how have we been able to maintain statistical power when the school sample size has gone down? First, the sample size has not gone down by very much – it is only slightly below the previous one. Second, our assumed average **total number of schools** per comparison (not per group) is **50** (there are three comparisons: control versus full, control versus partial, and partial versus full) and this is the same assumption we made when we carried out the previous (baseline) analyses. Third, we are looking at a **minimum** DES (MDES), and even with a slightly smaller sample size we are still in line with detecting that minimum. This is why the power analysis remained applicable across all three stages of the tests.

Modelling outcomes

When multi-level modelling was carried out, with pupil test scores as the outcome, no statistically significant differences among treatment groups were found at either post test 1 or post test 2 stages. In other words, the pupil-level ANCOVA (Analysis of Covariance) analyses did not show any statistically significance: at post test 1 - Effect Size (Full vs control) = + 0.02 ($p<0.70$); Effect Size (Partial vs control) = +0.05 ($p<0.26$). In addition, no statistically significant differences were found when year groups were used as a variable. The only statistically significant difference found was in relation to gender: girls performed better than boys ($p<0.001$).

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	**Coefficient**	**Standard error**	***t*-ratio**	**Approx. *d.f.***	***p*-value**
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	-0.024744	0.049365	-0.501	59	0.618
FULL, γ_{01}	0.018173	0.129287	0.141	59	0.889
PARTIAL, γ_{02}	-0.065592	0.107588	-0.610	59	0.544
ZP1, γ_{03}	0.507873	0.121507	4.180	59	<0.001
For GENDER slope, β_1 (Male=1, Female=0)					
INTRCPT2, γ_{10}	-0.263424	0.056910	-4.629	1323	<0.001
For Y3 slope, β_2					
INTRCPT2, γ_{20}	-0.410090	0.238290	-1.721	1323	0.085
For Y4 slope, β_3					
INTRCPT2, γ_{30}	0.380945	0.200451	1.900	1323	0.058

Final estimation of variance components

Random Effect	**Standard Deviation**	**Variance Component**	**d.f.**	χ^2	***p*-value**
INTRCPT1, u_0	0.35920	0.12902	59	251.26033	<0.001
level-1, r	0.87210	0.76055			

Statistics for current covariance components model

Deviance = 3672.213264

Number of estimated parameters = 2

We are a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health. We support the brightest minds in biomedical research and the medical humanities. Our breadth of support includes public engagement, education and the application of research to improve health. We are independent of both political and commercial interests.

Wellcome Trust
Gibbs Building215 Euston Road
London NW1 2BE, UK
T +44 (0)20 7611 8888
F +44 (0)20 7611 8545
E contact@wellcome.ac.uk
www.wellcome.ac.uk