Trial Evaluation Protocol
Primary Science Quality Mark
Evaluator (institution): RAND Europe
Principal investigator(s): Elena Rosa Brown

PROJECT TITLE	Primary Science Quality Mark
DEVELOPER (INSTITUTION)	PSQM, University of Hertfordshire
EVALUATOR (INSTITUTION)	RAND Europe
PRINCIPAL INVESTIGATOR(S)	Dr Alex Sutherland[1] (20 July 2018 – 14th June 2019) Dr Emma Disley (15 June 2019 – November 2019) Elena Rosa Brown (August 2019 – present)
PROTOCOL AUTHOR(S)	Dr Alex Sutherland Amelia Harshfield Dr Yulia Shenderovich Elena Rosa Brown Miriam Broeks
TRIAL DESIGN	Two-arm cluster randomised controlled trial with random allocation at school level
PUPIL AGE RANGE AND KEY STAGE	Whole-school intervention; focus of the evaluation on Year 5 (9-10 years old)
NUMBER OF SCHOOLS	140
NUMBER OF PUPILS	~3,500
PRIMARY OUTCOME	Pupil science attainment in study year 2[2]
SECONDARY OUTCOME	Pupil science attainment in study year 1, pupil attitudes to science and science teaching

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.2 [*latest*]		
1.1	June 2020	To capture changes to IPE and update project timelines in response to COVID-19
1.0 [original]	11 June 2019	

[1] Formerly RAND Europe, currently Behavioural Insights Team.
[2] Throughout the document, we refer to study year 1 (2019/20) and year 2 (2020/21).

List of abbreviations

CPD	Continuing Professional Development
EAL	English as additional language
EEF	Education Endowment Foundation
ES	Effect size
FIML	Full Information Maximum Likelihood
FSM	Free school meals
GDPR	General Data Protection Regulation
HSPC	Human Subjects Protection Committee
ICC	Intracluster correlation
ICO	Information Commissioner's Office
IDEA workshop	Intervention Delivery and Evaluation Analysis workshop
IPE	Implementation and process evaluation
ITT	Intention-to-treat
KS	Key Stage
NFER	National Foundation for Educational Research
MAR	Missing at random
MDES	Minimum detectable effect size
MI	Multiple Imputation
MoU	Memorandum of Understanding
NPD	National Pupil Database
PRU	Pupil Referral Unit
PSQM	Primary Science Quality Mark
RCT	Randomised Controlled Trial
SLT	School Leadership Team
TDTS	Thinking, Doing, Talking Science
TIMMS	Trends in International Mathematics and Science Study
STEM	Science, technology, engineering and mathematics
TIDieR	Template for Intervention Description and Replication
VLE	Virtual Learning Environment

Table of contents

Intervention

Primary Science Quality Mark (PSQM) was initiated in 2008 at the University of Hertfordshire to raise the profile of science in primary schools in England and promote professional development in science teaching and leadership.[3,4] PSQM is a developmental accreditation programme aiming to improve science education in primary schools through providing teachers and school science leaders with a framework for self-assessment, reflection and development as well as relevant training.

PSQM is delivered within hubs of schools (with a mean of 10 schools in a hub), supported by an experienced hub leader. Hub leaders have backgrounds such as Local Authority advisers, consultants, university lecturers and teachers who have achieved Primary Science Quality Marks in the past. Schools can work towards one of three Primary Science Quality Marks – PSQM, PSQM Gilt and PSQM Outreach. PSQM is for "schools which demonstrate how effective science leadership is *beginning* to have an impact on science teaching and learning across the school", whereas PSQM Gilt requires the demonstration of a "*sustained* impact", and PSQM Outreach is for schools that meet Gilt criteria and also impact science leadership and teaching in other schools.

Over the course of one academic year, PSQM involves the following activities (see Figure 1 below for the full logic model):

- Staff training, provided by the hub leader, completed over two full days or four half-days (topics: introduction to PSQM, creating and executing an action plan, and writing a reflective submission and collating appropriate supporting evidence).
- The subject leader works with colleagues across the school to audit existing provision in science and agree appropriate quality mark to work towards.
- The subject leader creates an action plan to develop aspects of science teaching, as specified in the PSQM framework and works with colleagues to implement it.
- Subject leaders are supported by the hub leader, with ongoing online mentoring provided via the PSQM Virtual Learning Environment (VLE), and access to resources such as the PSQM handbook and information on relevant Continuing Professional Development (CPD) offers.
- The subject leader collates and submits the evidence for the relevant PSQM, which is reviewed by a hub leader from another hub.
- Hub leader reviewers use PSQM evaluative criteria to consider whether a school has achieved the requirements to gain the chosen Primary Science Quality Mark.

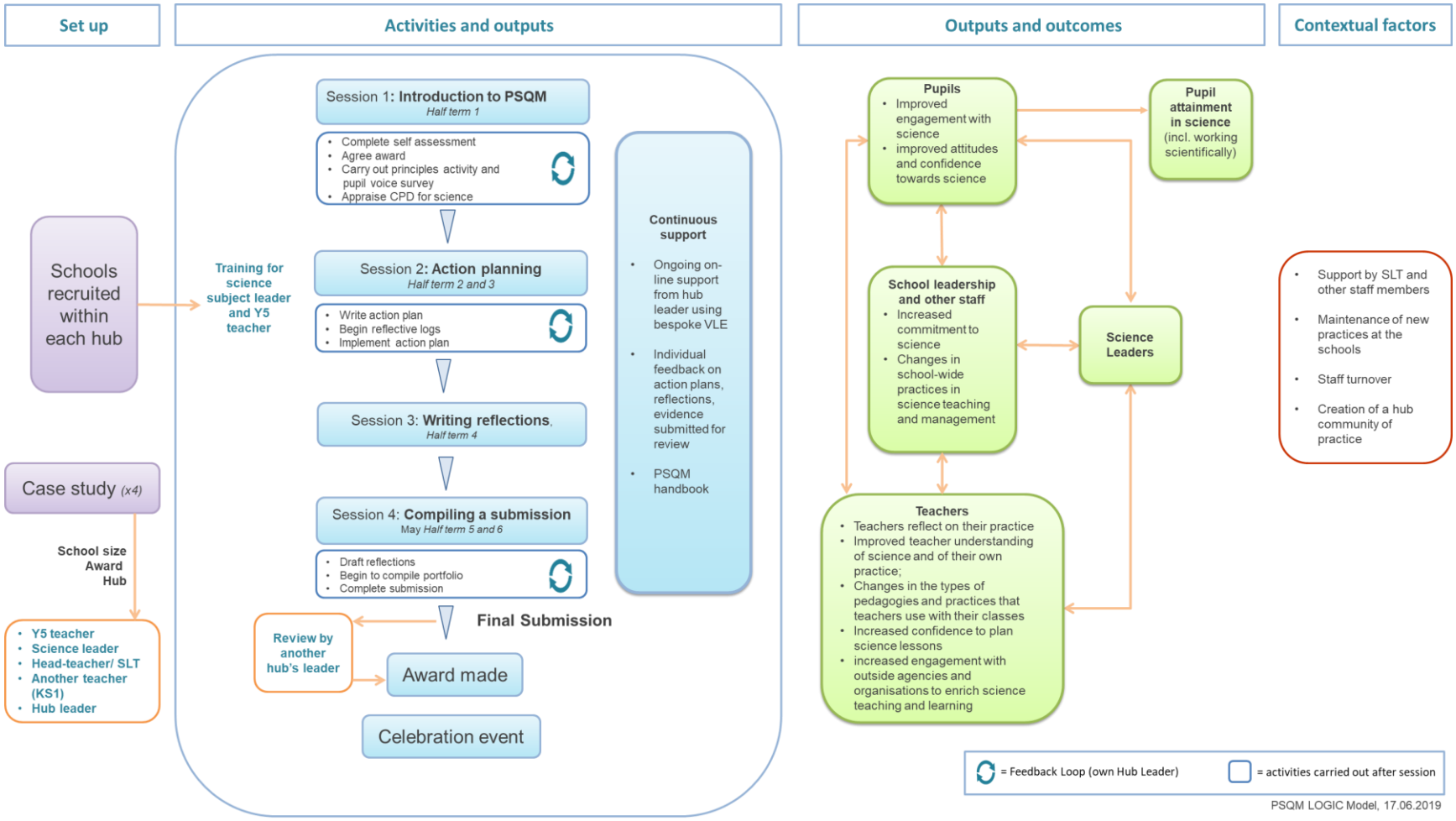
Awards are made to schools following an analysis of a series of documents that detail how the activities implemented during the intervention year have impacted on the science teaching and learning across the school and how the school meets the PSQM criteria. There are 13 PSQM criteria covering (1) primary school science leadership, (2) teaching (3) learning, and (4) wider opportunities Rather than the award itself being central, *the focus of the programme is on the process of self-assessment, reflection and development.*

All schools must complete the same self-evaluation and meet the same criteria, ensure that the subject leader (and another member of staff if possible) attend training, write and implement an action plan and submit common core documents. However, each school's action plan, implementation and final submission is relevant to its own context.

In the current trial, PSQM will be delivered in approximately 70 primary schools, with another approximate 70 schools assigned to the control arm. In the current evaluation, the programme will focus on the school's science subject leader and Year 5 teacher from each school (and a Key Stage (KS) 1 teacher, if the Y5 teacher is the subject leader).

[3] http://www.psqm.org.uk/what-is-psqm
http://www.psqm.org.uk/about-us
[4] http://www.psqm.org.uk/__data/assets/pdf_file/0010/123130/Primary-Science-May-2016-PSQM-update.pdf

Figure 1. PSQM logic model



Study rationale and background

Recent surveys of UK Science Subject Leaders and teachers in primary schools (CFE, 2017; CFE, 2019), including 902 science leaders and 1,010 teachers, suggested that science is often seen as less important compared to English and mathematics. Challenges reported in relation to science education in primary schools include lack of teaching time, lack of quality monitoring, limited access to science expertise, among others (Wellcome Trust, 2014; Ofsted, 2019).

PSQM is aimed at improving school-wide science teaching and raising the profile of science in UK primary schools through: (i) effective science leadership and (ii) supported school self-evaluation. PSQM is already widely used – more than 2,840 schools have previously completed the programme (11.8% of all UK primary schools) and more than 550 are currently engaged.[5] PSQM has also been endorsed by OFSTED (OFSTED, 2013), and is the only national award for science in English primary schools.[6]

Existing qualitative research suggests that PSQM can benefit schools in multiple ways, such as contributing to raising the profile of science in primary schools and providing schools with a framework and professional support for developing science leadership, teaching, and learning (White, et al., 2016). Previous evaluations of PSQM drew on interview, focus group, and survey data from participating science leaders and hub leaders. Participants reported that their perception was that PSQM improved the profile of science and quality of science teaching within schools and facilitated dissemination of relevant good practices between schools (White et al., 2016; White et al., 2015).

However, there is no robust experimental evidence yet on whether PSQM accreditation leads to improvements in pupil outcomes in science or related subjects. The current study aims to produce rigorous evidence on PSQM's efficacy in relation to pupil outcomes in science.

Previous evidence is limited regarding the impact of accreditation programmes in primary and secondary education. However, existing literature suggests that accreditation programmes in higher educational institutions can improve the quality of teaching (Hanbury et al., 2008; Volkwein et al., 2006; Blouin et al., 2018). There is also evidence from survey data that accreditation translates into better outcomes for university students (Volkwein et al., 2006). If the criteria that schools must meet in order to gain accreditation lead to improved pedagogical methods and, consequently, improved learning by students, this intervention should lead to improved attainment.

Furthermore, the active and collaborative style of professional development that PSQM draws on has been linked to positive effects on instructional practice and student outcomes (Opfer, 2016; Darling-Hammond et al., 2017; Gore et al., 2017). Nevertheless, CPD programmes that are active and collaborative do not always lead to improvements in pupil outcomes (e.g. Garet, 2011; 2016; Sims and Fletcher-Wood, 2018).

The PSQM programme is led by The University of Hertfordshire and will be independently evaluated by RAND Europe. The study is funded by the Education Endowment Foundation (EEF) and Wellcome Trust.

[5] http://www.psqm.org.uk/about-us
[6] http://www.psqm.org.uk/__data/assets/pdf_file/0006/78405/PSQM-flyer-July-2017.pdf

Impact Evaluation

Research questions

The impact evaluation is designed to investigate the following research hypotheses[7]:

Hypothesis 1: Year 5 pupils in randomly allocated primary schools participating in PSQM (intervention schools) will have higher levels of science attainment than the pupils in the comparison schools one year following the end of PSQM implementation, 2020/21 (Summer 2021; primary outcome).

Hypothesis 2: Year 5 pupils in primary schools participating in PSQM (intervention schools) will report higher levels of enjoying science than the pupils in the comparison schools in 2020/21 (Summer 2021; secondary outcomes).

Design

Trial type and number of arms		Two arm stratified, cluster-randomised controlled trial, randomised at the school level
Unit of randomisation		School
Stratification variables (if applicable)		Region (hub) School size (single- versus multiple-form entry)
Primary outcome	variable	Pupil science attainment (Year 5 pupils in summer 2021)
	measure (instrument, scale)	Hanley 2015 (potentially modified to ensure fit with the National Curriculum), science assessment with scores ranging from 0 to 41 points
Secondary outcome(s)	variable(s)	Pupil attitudes to science and science teaching
	measure(s) (instrument, scale)	Trends in International Mathematics and Science Study (TIMMS)

The PSQM evaluation will be a two-group parallel, stratified, cluster-randomised trial, with school as the unit of randomisation. To ensure comparability of schools in the intervention arm and the control arm ('exchangeability', see Oakes, 2013), we will randomise schools within hubs, which will serve to balance the study arms on geographical location and, therefore, any regional differences.[8]

During the recruitment period (2018-19 academic year), schools are asked to nominate one Year 5 teacher (in case there are multiple Year 5 classes) to participate in PSQM. The class of this teacher is considered the focal class for the evaluation, assessed in the summer 2020 following implementation. In Summer 2021, the Year 5 class taught by the same teacher will be assessed. If the teacher has left or moved to another Year, we will assess the Year 5 class or randomly select another Year 5 class (if there are more than one Year 5 classes).

To minimise the burden on pupils and schools, the evaluation will use administrative data for baseline, with schools providing pupil identifiers, which will be linked to the National Pupil Database (NPD). After

[7] The initial trial design included a third hypothesis, namely: Year 5 pupils in primary schools participating in PSQM (intervention schools) will have higher levels of science attainment than the pupils in the comparison schools at the end of the school year when the intervention takes place, 2019/20 (Summer 2020; secondary outcome). Due to testing in the summer 2020 being cancelled due to COVID-19, it is no longer possible to test this hypothesis and has therefore been removed.
[8] That is – if one were to swap the intervention and control groups the results from the trial should be the same.

schools have been recruited and the pupil and teacher information collected, the Evaluation Team will randomise schools to one of two arms: intervention or control.

Intervention schools will not be charged to take part in the PSQM programme and will receive a payment of £1,500 towards teaching cover and £120 towards travel costs. Control schools will not be allowed to participate in PSQM while the study is running but they will receive a payment of £1,500 on completion of the trial.

Randomisation

Randomisation will be conducted in Stata by the Evaluation Team's Primary Investigator. Hub will be the main stratifying variable, with around 16 hubs expected to be recruited. In addition, we plan to stratify on school size (single-entry versus multiple-entry school), as reported by the school. The trial allocation will be recorded and communicated to the implementation team and the EEF in a password protected Excel file to prevent editing. Initial outcome analyses will be conducted blind to allocation.

Baseline equivalence will be examined based on the initial randomisation. A well-conducted randomisation will, in expectation, yield groups that are equivalent at baseline (Glennerster & Takavarasha, 2013). Because schools are randomly allocated to the control and intervention conditions, any imbalance at baseline will have occurred by chance. To assess imbalance at baseline, we will compare groups at school and pupil levels, by means of cross-tabulations and histograms that assess the distribution of each characteristic within the control and intervention groups.

Participants

SCHOOLS

Schools are recruited by the PSQM team and PSQM hub leaders, based on the following eligibility criteria:

Inclusion criteria:

- The school cannot have received a PSQM award in the last 3 years (i.e., a school has not participated in PSQM in 2017, 2018 or 2019).
- The school must be a state primary, junior or all-through school.
- Schools with mixed Year 5/6 or another combination are eligible if they have Year 5 pupils taught separately by one teacher for science.

Exclusion criteria:

- Infant or first schools, private schools, special schools, Pupil Referral Units (PRUs) or middle schools are not eligible.

The following areas were included in the recruitment:

- Aylesbury Vale
- Barnsley and Kirklees
- Bracknell and Slough
- Cambridgeshire (East)
- Cannock
- Chorley
- Crewe and Nantwich
- Merton
- Cumbria
- Devon (North)
- Essex
- Isle of Wight
- Loughborough
- Newent
- North Yorkshire
- Oxford & Banbury
- Portsmouth
- Ross-on-Wye
- Suffolk coastal
- Tewksbury
- Thanet and Medway
- Waltham Forest
- Warrington

PUPILS

There are no inclusion/exclusion criteria based on pupil characteristics as PSQM is a universal intervention. To minimise burden on schools, pupils enrolled at the time of school recruitment in 2019 are included in the study, but pupils who join the schools at a later time will not be included in the evaluation as this would require additional information collected from schools.

Minimum Detectable Effect Size (MDES) calculations

Table 1. Statistical power calculations

		Main effect
Minimum Detectable Effect Size (MDES)		**0.197**
Pre-test/ post-test correlations	level 1 (pupil)	0.63
	level 2 (class)	NA
	level 3 (school)	0
Intracluster correlations (ICCs)	level 2 (class)	NA
	level 3 (school)	.15
Alpha		0.05
Power		0.8
One-sided or two-sided?		Two
Average cluster size[9]		25
Number of schools	Intervention	70
	Control	70
	Total	140
Number of pupils	Intervention	1,750
	Control	1,750
	Total	3,500

Power and minimum detectable effect size (MDES) calculations were performed using the PowerUp tool for main effects (Dong & Maynard, 2013) and moderators (Spybrook, Kelcey, & Dong, 2016; Dong, et al., 2017). Based on EEF guidelines (EEF, 2018) and on a recent evaluation working with science outcomes in this age group (Kitmitto 2018)[10], the amount of variation explained by covariates for 140 schools with an average of 25 pupils each, is assumed to be 0.40 (equivalent to correlation of 0.63) for level 1 (pupils) and 0.00 for level 2 (schools). The efficacy evaluation of Thinking, Doing, Talking Science (TDTS), which used the same primary outcome (Hanley et al., 2015) reported an intracluster correlation (ICC) of 0.15 in the analyses. With one class per school included in the evaluation, we assume an average cluster size of 25 pupils. We also assume an alpha of 5% and an intended 80% power to detect effects. We use two-level clustered designs, assuming a continuous, normally distributed (Gaussian) outcome.

[9] We have set the average class size to 25, but acknowledge that there may be variation across schools where some classes are smaller with less than 20 pupils, and others are larger with up to 30 pupils.

[10] The effectiveness evaluation of TDTS (Kitmitto, 2018) found variance explained at Level 1 to be 0.40 for the same primary outcome as in the current trial and KS1 reading/writing and mathematics as baseline, so we expect the current trial to have at least the same variance explain as a minimum.

Using the parameters above and with equal allocation to intervention and control the MDES is 0.197 (Column A). We believe it would be important to power to 0.2 even though this is an efficacy trial because the universal nature of the intervention is likely to result in comparatively smaller effect size.

Based on EEF's guidance[11], we focus on a moderator effect defined as a statistical interaction of intervention and moderator variables. Based on the average number of free school meals (FSM) pupils in UK primary schools – 14% in 2018 - we assume 4 FSM pupils per class.[12] However, PSQM recruitment for the trial focused on high-FSM areas, so the actual number may be higher. Using the same assumptions as the main analysis, MDES difference regarding Cohen's d is 0.251 (95% CI 0.059; 0.331).

Baseline measures: The evaluation will use pupils' KS1 mathematics, reading and writing data collected in Year 2 as baseline data to assess baseline equivalence of the intervention and control groups during the randomisation process of the schools. These data will also be used as covariate(s) in outcome analyses.

Outcome measures

Primary outcome

We propose using an independent science test at post-test, administered and marked by a third-party, the National Foundation for Educational Research (NFER) for outcome testing. This approach allows for blinding to allocation, as we can supply a list of schools to the assessors without revealing allocation. KS2 science is teacher assessed and would, therefore, bring the problem of biased measurement/non-blinding.

We plan to use the test on knowledge, thinking and reasoning in science used in the EEF evaluations of TDTS (Hanley et al., 2015). This test was compiled from questions developed by Terry Russell and Linda McGuigan for an unrelated Randomised Control Trial (RCT) funded by the Wellcome Trust and covers a range of topics in biology, chemistry and physics. It includes process/inquiry-based, concept-based; and open-ended conceptually-based questions. The test is currently under external review by a team at York University, commissioned by the EEF, to ensure compatibility with the current National Curriculum. The team is redesigning the instrument in four phases during 2020 involving piloting and validation. In January 2020 it was piloted in two schools with 24 children (phase 1) and in February in 22 schools with 958 pupils (phase 2). Phase 3 is planned for October 2020 to test the psychometric properties of the test in 14 schools. It will involve piloting a version of the test made up of 15 questions and lasting approximately 45 minutes. Phase 4 (November 2020) will assess the test-retest reliability of the instrument. Any changes to the outcome(s) will be conveyed through updated versions of this protocol.

As whole-school changes take time, we will evaluate the effect of the intervention on pupil science attainment and attitudes following the 2019/20 implementation year with the second Year 5 cohort, in 2020/21 (labelled Cohort B) in all schools, namely in both intervention and control groups (see Table 2). Initial plans were to also evaluate the short-term effect of the intervention with Cohort A in 2019/20. However, due to COVID-19 related lockdown and social distancing measures made pupil testing impossible at the end of the 2019/20 school year.

11 https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

12 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/719226/Schools_Pupils_and_their_Characteristics_2018_Main_Text.pdf

Table 2. Key data collection points over time

	2018/19	2019/20	2020/21
Study activities	Programme recruitment	PSQM implementation in the intervention schools; business as usual in control schools	Business as usual Pupil Cohort B tested
Outcome measures		None*	Science attainment among Year 5 pupils in the nominated teacher's class, or, if the teacher not in Year 5, the year 5 class or if there is more than one Year 5 class a randomly selected Year 5 class *(Primary outcome).* Attitudes to science and science skills among Year 5 pupils: - Science enquiry skills - Enjoyment of science *(Secondary outcomes).*
Process evaluation measures	See Table 3	Headteachers, nominated Year 5 teachers and subject leaders surveyed. Headteachers, teachers and subject leaders from case study schools interviewed about PSQM processes and science teaching. Hub leader survey Hub Leaders from selected case study schools interviewed. See Table 3 for full list.	Headteachers, nominated Year 5 teachers (working with Cohort B), and subject leaders surveyed. Headteachers, teachers and subject leaders from case study schools interviewed in follow-up interviews about PSQM processes and science teaching in year 2 of the trial. See Table 3 for full list.

* Initial plans were to administer the science attainment test and the 'attitudes to science and science skills' test among Year 5 pupils in the nominated class at the end of 2019/20, both as secondary outcomes. However, school closures resulting from COVID-19 led to the cancelation of testing in the summer of 2020.
** **All outcomes will be collected by NFER, blinded to allocation. Pupil attainment will be marked by NFER, while pupil attitudes measures will be scanned (entered) and summarised by the RAND Europe team.**

Secondary outcomes

To assess changes in pupils' attitudes towards science, we suggest carrying a post-test survey at the same time as the primary outcome assessment. The attitudinal measure at post-test will also be administered by NFER using paper forms. The attitudinal measures will be compiled in machine-readable forms, to allow scanning, data entry and scoring by RAND Europe.

Enjoyment of science, confidence in science and engaging teaching in science will be measured using the 'enjoyment of science' subscale adapted from the Trends in International Mathematics and Science Study (TIMMS) Grade 4 surveys from TIMMS 2015.[13]

[13] https://timssandpirls.bc.edu/publications/timss/2015-methods.html

If possible, science enquiry skills will be captured from relevant items in the science attainment test by Hanley and colleagues.

Analysis plan

The primary outcome for the second wave of Year 5 pupils (Cohort B) will be science attainment as measured by the science test. Intervention and control arms will be compared in terms of the difference in means between groups at follow-up, conditional on baseline measures (KS1 mathematics, reading and writing) and stratification variables (area and school size).

The unit of analysis here will be pupils. There is an ongoing discussion about how 'best' to analyse results from RCTs that involve clustered data. One approach, 'analyse how you randomise' (Senn, 2004), suggests that one should explicitly account for clustering via multilevel models (AKA 'random effects'). This approach assumes that the schools in the study are a random sample of all schools – which is often a source of contention – but one benefit of this approach is being able to explicitly partition variance and more flexibly handle complex variation within schools (Snijders and Bosker, 2012). Our approach will be to conduct sensitivity analyses to assess results against different model specifications. These will be detailed in the Statistical Analysis Plan. The general equation for the multilevel model is given below as Eq.(1):

$$y_{ij} = \propto + X_{ij}\beta + Z_j b_j + \delta PSQM_j + u_{ij} + u_j \qquad i = 1..N, j = 1..M, \quad (1)$$

where y_{ij} denotes the pupil level outcome; i and j denote pupil and school indexes, respectively; X_{ij} is the $1 \times k$ vector of individual characteristics that include the KS1 measures as a pre-test;[14] Z_j is a vector of the stratification variables mentioned above (hub region and school size); $PSQM_j$ is a dummy variable denoting intervention /control group at the school level; β and δ are the $k \times 1$ and 1×1 vectors of regression coefficients; u_{ij} is the pupil-level error term; and u_j is the school-level error term. The coefficient δ will constitute the main result of the trial.

The outcome analysis will be on an intention-to-treat (ITT) basis. Once randomised, schools and participants will be analysed according to the allocation of the school regardless of whether the school complied with the intervention or not. It is important to note that cluster-randomised designs mean that both school and pupil level attrition may be possible post-randomisation, with subsequent implications for analysis (see Schochet and Chiang, 2011).[15] The ITT approach is inherently conservative as it captures the averaged effect of *offering* the intervention.

Our approach would be to adhere to the ITT analysis in the event of pupils migrating between intervention and control schools after randomisation. Pupils joining schools after the new school year had begun would be excluded from the evaluation.

Baseline data

The baseline pupil measure from the NPD will be used as a continuous variable. KS1 mathematics, reading and writing will be collected approximately one to two years before randomisation (for Cohort B pupils). These pupil data will be obtained from the NPD, based on the lists of pupils. Pupil information will be provided by all trial schools during recruitment and before schools are randomly allocated to control or intervention conditions. NPD baseline data will be matched to the science attainment scores for each pupil.

[14] Assuming that the KS1 measures are not too highly correlated to be included in the same model e.g. if the correlation between measures is *r* >= .7 we would include only one measure (e.g. KS1 reading).

[15] While not widely known or reported, random effects models may yield biased estimates of ITT in cluster randomised trials under certain conditions when there is individual level noncompliance. Thus, it is critical to minimise individual level noncompliance and to include adequate covariates to reduce between-cluster variance. See for example Jo et al. (2008).

Effect size (ES)

We will use the effect sizes for cluster-randomised trials given in the EEF evaluator guidance – an example, adapted from Hedges (2007) is given below:

$$ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\sigma_S^2 + \sigma_{error}^2}}$$

Where $(\bar{Y}_T - \bar{Y}_C)_{adjusted}$ is the mean difference between intervention groups adjusted for baseline characteristics and $\sqrt{\sigma_S^2 + \sigma_{error}^2}$ is an estimate of the population standard deviation (variance). The ES therefore represents the proportion of the population standard deviation attributable to the intervention (Hutchison and Styles, 2010). The exact effect size used will depend on whether there are equal or unequal sample sizes in trial arms.

Same approach for primary and secondary outcomes

Moderator analyses

Two moderators will be examined to explore intervention heterogeneity:

1) EverFSM
2) Gender - motivated by the gender gap in Science, technology, engineering and mathematics (STEM) careers in adult population.

Missing Data

Attrition across trial arms will be explored as a basic step to assess bias (Higgins et al., 2011). To gauge systematic differences between those who drop out and those who do not – and whether factors should be included in analysis – we would model missingness at follow-up as a function of baseline covariates, including intervention. For item non-response, the extent of missingness may in part determine the analytical approach.

For less than 5% missingness overall a complete-case analysis should suffice, regardless of the missingness mechanism (EEF, 2018). Our default would be to check results using approaches that account for missingness that rely on the weaker missing at random (MAR) assumption, building the MAR conditioning variables from our initial work predicting missingness. If there was systematic missingness of predictor variables, for example, we would explore options for using Full Information Maximum Likelihood (FIML) and/or multiple imputation (MI) (EEF, 2018; for a discussion of FIML vs MI see Allison, 2012).

Implementation and process evaluation

The process evaluation will address the following questions:

- Was the intervention implemented with fidelity for the intervention schools?
- What was practice as usual in the control schools?
- What appear to be the necessary conditions for success of the intervention?
- What were the barriers to delivery?

We have developed a multi-stage mixed-methods Implementation and Process Evaluation (IPE) data collection plan. We will collect data through monitoring data, surveys for all schools in the trail (intervention and control), in addition to a documentary review, interviews and school visits (to observe science displays) for selected case study schools (see Table 3). Upon detailed review of PSQM's logic model it was felt that visits to schools were a way to observe some of the key PSQM intended changes.

Pre-intervention

Non-enrolment and drop-outs

While it is not possible to identify the characteristics of the schools that do not participate in the trial, the PSQM team will monitor the contact with schools and the general sign-up rate in order to get a sense of non-enrolment. If a school drops out from the programme, the hub leader will be responsible for notifying the PSQM team with accompanying reasons for why the drop-out occurred.

DURING INTERVENTION PHASE

Motivations for joining the study

The baseline headteacher online surveys (September – October 2019) will aim to examine the motivations for joining the trial and the current practice related to science teaching. It is important to note that two other recent research efforts (CFE 2017; 2019) also describe the existing practices regarding science teaching in primary schools in England.

Completion of intervention activities by intervention schools

We will assess the attendance rate at PSQM training sessions. Attendance of trainings by teachers is mandatory and will be tracked by the PSQM team through attendance logs. We do not anticipate non-attendance to be a substantial problem given that by signing-up, schools have committed to send teachers to training sessions – however, attendance is still an important metric to capture. PSQM will put together milestones to measure successful programme implementation and participant involvement. Other data collected by PSQM to monitor implementation fidelity includes logging onto VLE, uploading action plans, the upload of core documents, reflections and submission. Implementation fidelity will be analysed for all intervention schools through a compliance measure (see Study Analysis Plan). In addition, an in-depth analysis of the described documents will be conducted for the selected case study schools.

Based on information provided by the PSQM team, we will also report on the number of schools that are successful in gaining the quality mark they aimed for, and the numbers of cases when a submission was sent for a second review, a school was asked to submit additional evidence to get the quality mark and/or when a school had a deferral/extension.

POST-INTERVENTION

Exploring programme implementation and changes in practice

Online surveys

Online surveys for headteachers, teachers, subject leaders and hub-leaders will be rolled out as part of the IPE activities.

One of the aims of the post-intervention headteacher surveys (in December 2020 and again in June 2021) will be to capture any potential changes in science practice (in both intervention and control schools). Any such identified changes will be highlighted in the evaluation report.

There is a possibility that staff members from various intervention and control schools discuss the intervention amongst themselves, thereby potentially leading to chances in science practices in the control schools (phenomenon known as “spillover” or “contamination”). However, given that this is a school-based intervention, we do not anticipate the likelihood of this occurring to be large.

In December 2020 and again in June 2021 online surveys will be distributed to two staff members per school (the subject leader and the nominated classroom teacher).[16] The focus of the surveys will be on usual practice, attitudes, perceptions and science-related activities in the classroom. The majority of

[16] Initial plans were to administer these surveys on paper at the same time as pupils would be tested. Due to COVID-19 trial adaptation it was decided to administer these surveys online.

survey items will be the same in both surveys for control and intervention schools, examining practices and attitudes around teaching of science and science-specific CPD activities. The questions will be based on relevant expected outcomes as defined by PSQM, and will draw from the Wellcome State of the Nation surveys for teachers and subject leaders where appropriate (CFE, 2017). The first survey will also ask about how COVID-19 affected running science-related activities. In addition, in the intervention arm, intervention-specific questions will be included based on the expected intervention outcomes outlined in the logic model. Descriptive quantitative analyses will be used to analyse survey data using Stata.

Hub leader surveys will focus on their experiences of working with schools and any barriers and facilitators to implementation. We expect that it will take no more than 10-15 minutes to complete the online surveys. The text for the survey will be prepared, compiled and distributed by RAND Europe .

Case studies

- *Interviews with staff in case study schools*

In addition, five schools among those assigned to the intervention group representing a diverse set of characteristics, will be approached for in-depth case studies. These will involve interviews with school teaching staff and headteachers, as well as– where possible – school governors. There will be two rounds of interviews, one in December 2020, another in June 2021. For the latter, a member of the Evaluation Team will conduct school visits to conduct the interviews in person and to observe whether there are any science boards displayed around the school and document this. This will also increase the chances to interview other relevant school stakeholders such as governors.

These interviews will allow the Evaluation Team to gain a more in-depth understanding of what PSQM involves in practice for participating teachers, subject leaders, and schools, and explore the mechanisms of change as a result of the intervention. This information is particularly important to understand what other schools need to do if they chose to participate in PSQM later. Bigger schools may find it more challenging to disseminate the impact. The dimensions that will be taken into account for sampling case study schools will be the type of award the school is working towards, whether they are single or multi-form, and location (hub).

NVivo software will be used to facilitate the development of a coding matrix using the transcripts from these interviews, following framework principles, with built-in flexibility to allow identification of anticipated and emergent themes.

- *Documentary review*

Documentation obtained through PSQM's VLE and provided by schools will be reviewed for case study schools only. Selected schools will be asked to provide documentation that captures their science-related activities such as, school development plans, school science policy plans (if available), Ofsted reports, lesson observation notes, feedback on school improvement plans (SIPs), reports for and communications with governors, and letters to parents. We will also seek to obtain pictures of science displays around the school.

Examining continuity of Year 5 teachers

We would expect programme effects to be strongest for Year 5 pupils in those schools where Year 5 teachers trained in PSQM continue teaching the next Year 5 cohort (Cohort B). To examine this, we will descriptively compare programme outcomes for those schools where the same teachers are working in Year 5 in 2020/21. The information on continuity will be based on self-report by teacher in Year 2 of the trial as part of their surveys.

Table 3 Overview of IPE data collection

Data type		Participant	When	Who collects the data	Topics
Online Surveys	**Headteacher survey 1**	Headteacher/SLT	Year 1, September 2019	RAND to design, and share with schools	- Experience with other trials/research school status (all schools) - Usual practices around teaching science (all schools)
	Headteacher survey 2	Headteacher/SLT	Year 1, (postponed to start) December 2020	RAND to design, PSQM to share with schools	- Usual practices around teaching science (all schools) - Intervention costs (intervention schools only)
	Teacher survey 1	Y5 Teachers (teacher selected for PSQM)	Year 1, (postponed to start) December 2020	RAND to design, PSQM to share with schools	- Experience with intervention activities (intervention schools only) - School's commitment to science (all schools) - Teacher's background (highest education in science, years teaching, years in the school) (all schools)
	Science subject leader survey 1	Science subject leaders	Year 1, (postponed to start) December 2020	RAND to design, PSQM to share with schools	- Experience with intervention activities (intervention schools only) - School's commitment to science (all schools) - Leader's background (years as subject leader) (all schools)
	Hub leader survey 1	Hub leader	Year 1, (postponed to start) December 2020	RAND to design, PSQM to share with schools	- Interactions with the schools, perceived level of school engagement, perceived barriers and enablers (intervention schools only)
	Headteacher survey 3	Headteacher/SLT	Year 2, June – July 2021	RAND to design, PSQM to share with schools	- Usual practices around teaching science (all schools) - Sustainability of changes related to PSQM (intervention schools only)
	Teacher survey 2	Nominated Y5 teachers	Year 2, June – July 2021	RAND to design, PSQM to share with schools	- School's commitment to science (all schools) - Whether Y5 teacher who took part in PSQM continued teaching the Y5 or not (interventions schools only) - Sustainability of changes related to PSQM (interventions schools only)
	Science subject leader survey 2	Science subject leaders			

Data type		Participant	When	Who collects the data	Topics
Case studies	**Interviews**	Teacher, headteacher, governor (3 people per school)	Year 1, (postponed to) December 2020 and Year 2, June – July 2021	RAND to design/conduct, PSQM to participate in selection of schools	(Case study activities applicable to interventions schools only) - Experience with intervention activities - Usual practices around teaching science - School's commitment to science - Sustainability of changes related to PSQM
		Hub leader	Year 1, (postponed to) December 2020	RAND to design/conduct, PSQM to help contact participants	- Experience delivering programme and perception of school engagement (barriers and enablers)
	Documentary review	Schools	Year 2, June – July 2021	PSQM provides VLE documentation to RAND. RAND requests other relevant documentation from schools	- Assess evidence of science presence/relevance in school plans and communications
Data on previous rounds of PSQM		Schools	Set-up year (September 2021)	PSQM to share with RAND	- % FSM - single/multiple entry To compare with RCT schools (all schools)
Monitoring data from PSQM		Schools	From trial start up until PSQM submission due February 2021	PSQM to share with RAND	- Non enrolment numbers/reasons (all schools) - Post-randomisation drop-out/reasons for drop-out (all schools) - Training attendance logs (intervention schools only) - School task completion logs (intervention schools only)

Compliance measure

To enable a non-compliance analysis, compliance will be defined at the school level, based on completion of programme activities, as recorded by the PSQM team. This will be specified in the Statistical Analysis Plan.

Cost evaluation

Cost data will be gathered through online surveys, as well as through the interviews in the implementation and process evaluation (see above). Questions will be targeted at assessing any pre-requisite costs (such as training costs and materials) and any direct and marginal costs directly attributable to schools' participation in the intervention (printing, staff time, cover, etc.). We will use this information to estimate cost per-pupil, following EEF guidelines (EEF, 2015).

The main costs of the intervention relate to training, materials, and the time of teachers and subject leaders to complete the programme activities. To calculate the cost of training and materials the Evaluation Team rely on data provided by the Delivery Team. RAND will also take into account the cost of the time of hub leaders, headteachers, teachers and other staff in delivering the programme.

We acknowledge that in the RCT, schools in the intervention arm of the trial will have £1,500 paid toward these costs and a further £120 for travel time and will take this into account.

We will use the information on direct and indirect costs to estimate cost per-pupil, following EEF guidelines (EEF, 2018).

Ethics and registration

The trial has been registered on the ISRCTN registry, which stands for 'International Standard Randomised Controlled Trial Number' and is used to describe RCTs and efficacy trials at inception. The trial has been assigned an ID registration number: ISRCTN50771738.

The ethics and registration processes are in accordance with the ethics policies adopted by RAND Europe. The evaluation is currently reviewed by RAND U.S. Human Subjects Protection Committee (HSPC).

Parents or legal guardians act as decision-makers for individual pupils. This is because the intervention will be delivered during the school day, where schools act in loco parentis, and the intervention does not substantially differ from standard practice in schools. Prior to pupil data being sent to the Delivery Team, parents will be sent information and withdrawal forms by the school and have the opportunity to return these. The parental information sheets and withdrawal forms will be sent out to parents by the schools after the school representative sign the Memorandum of Understanding (MoU) describing what is involved in the trial. Parents can withdraw their children at any time from the research, but the initial withdrawal forms can be returned by parents within two weeks.

If participants choose to withdraw their children from the study later on, their data will not be collected or will be deleted, as appropriate (see Privacy Notice at http://redocuments.org/PSQM/Privacy_Notice_Parents.pdf).

RAND Europe will collect consent forms for school staff, governors and parents who will volunteer to participate in an interview. Furthermore, the cover page for each survey will contain a privacy notice for respondents. It will inform respondents that participation in the survey is entirely voluntary.

None of the Evaluation Team has any conflicts of interest and all members of the study team have approved this protocol prior to publication.

Data protection

RAND will obtain personal data from schools and pupils as data controller. Basic pupil information will be obtained on the basis of legitimate interests from schools pursuant to brief data sharing undertakings or agreements with each school recruited. RAND shall obtain pupil baseline and outcome data from its subcontractor (e.g., NFER), who will act as a processor pursuant to appropriate data sharing terms in it subcontract. Data obtained by NFER is expected to be on the basis of legitimate interests and pupils and parents shall be provided with age-appropriate fair processing privacy notices that explain the use, storage and secure handling of the data.

Data sharing agreements between the parties will outline in detail how and which data will be securely shared between them using the secure platform "Syncplicity". Data will only be saved on General Data Protection Regulation (GDPR) compliant, secure servers inside the EEA or UK. All processes will be handled in accordance with RAND's Data Protection Policy. RAND is registered with Information Commissioner's Office (ICO), registration number Z6947026 and is certified for adhering to ISO 9001:2015 quality management practices. In order to stratify the sample and adequately evaluate the intervention as outlined in this proposal, it is necessary to process special categories of data, namely FSM status of pupils. RAND Europe considers this endeavour to fall under GDPR, Chapter 2, Article 9, Paragraphs 2d) and 2g).

Personnel

DELIVERY TEAM: PSQM (UNIVERSITY OF HERTFORDSHIRE)

Project Leader and PSQM Director: Associate Professor Jane Turner (University of Hertfordshire)

PSQM Deputy Director: Helen Sizer (University of Hertfordshire)

PSQM team: Claire Warren (University of Hertfordshire)

EVALUATION TEAM: RAND EUROPE

Overall Project & Evaluation Lead: Elena Rosa Brown (took over Dr Emma Disley in November 2019; previous project lead was Dr Alex Sutherland until June 2019) (RAND Europe).

Project Manager: Miriam Broeks (took over from Amelia Harshfield in March 2020; previous project manager was Dr Yulia Shenderovich until April 2019) (RAND Europe)

Core fieldwork and analysis team: Miriam Broeks, Sashka Dimova Amelia Harshfield (all RAND Europe)

Risks

Risk	Assessment	Mitigation strategy
Recruitment failure	Likelihood: Low Impact: High	Remain in dialogue with the PSQM Delivery Team over any recruitment issues. Provide letters for schools explaining the research process. Seek support from the EEF to encourage recruitment.
Attrition	Likelihood: Moderate Impact: Moderate to high	Clear information about expectations and requirements provided to participating schools. MoU to be signed with participating schools. Attrition to be monitored and reported according to CONSORT guidelines (Campbell et al., 2010). Schools in control group will receive a proportion of their payment for participating in the trial after outcomes testing has been completed in year one and the second amount at the end of year two. This is an incentive to remain in the trial.
Different rates of attrition from control and intervention groups	Likelihood: Low Impact: Moderate	There is a risk that schools in the intervention group may face an extra burden in terms of time and resources to deliver the programme. This can be mitigated by regular liaison with hub leaders and schools to secure continued engagement in the trial. There is a risk that control group schools may decide to withdraw from the trial because they wish to take part in PSQM and signed up to the trial in the hope they would be in the intervention group Schools would have agreed to the terms of the MoUs, which include the commitment for data to be collected at various stages.
Missing data	Likelihood: Moderate Impact: Moderate	To limit the amount of missing data screening, testing will happen in an extended period (approximately a month).
Pupil mobility	Likelihood: Moderate Impact: Low	Pupils who migrate to non-study schools will be excluded from the analysis as these pupils will be tested with external tests. In the event that mobility to non-study schools exceeds 10% on average across all schools, then the evaluators will discuss with the EEF the possibility of additional funding to collect this information.
Low implementation fidelity	Likelihood: Low to moderate Impact: Moderate	Process evaluation to monitor and document fidelity of implementation. Remain in dialogue with the PSQM Delivery Team on finding solutions.
Cross-contamination	Likelihood: Low Impact: High	Clear instructions will be provided to participants about the trial to avoid contamination.
Evaluation team members absence or turn-over	Likelihood: Moderate Impact: Low	All RAND staff have a three month notice period to allow sufficient time for handover. The team can be supplemented by researchers with experience in evaluation from the larger RAND Europe pool.
Low response rates for online surveys	Likelihood: Moderate Impact: Moderate	Online surveys to be kept to a maximum of 5-15 minutes long. Respondents given the opportunity to complete survey online on multiple occasions if required. Sufficient data collection window given with real-time monitoring of response rates to allow for reminders to be targeted. This may be a more significant problem in the control group. To address this, control schools will receive a payment of £1,500 on completion of the study.

Lack of coordination across the EEF (funders), RAND Europe (evaluators) and the PSQM team (delivery team)	Likelihood: Moderate Impact: Moderate	Teams to attend initial meetings and agree on roles and responsibilities at the outset. Regular updates to be provided to the lead evaluators. Regular contact between senior team from each organisation.
Further disruptions to trial activities due to COVID-19 (or alike) pandemic	Likelihood: Moderate Impact: High	A new outbreak of the COVID-19 pandemic could further delay evaluation activities. Teams will maintain regular contact to coordinate and decide on any strategies to mitigate potential risks to the evaluation plans.

Timeline

Dates	Activity	Staff responsible/ leading
October 2018	IDEA workshop	RAND Europe
January – April 2019	Recruiting schools and teachers	University of Hertfordshire
January – April 2019	Opt out forms to be sent to parents	Schools
March – May 2019	Collection of pupil information	Schools
May - June 2019	School and pupil information to be collected sent to RAND	University of Hertfordshire
May – June 2019	Randomisation	RAND Europe
September – October 2019	Baseline survey of headteachers all schools	RAND Europe
September 2020	Completion of Statistical Analysis Plan	RAND Europe
December 2020 – February 2021*	IPE surveys and interviews	RAND Europe
September 2019 – February 2021*	Programme implementation	University of Hertfordshire
February – March 2021*	Compilation of CPD attendance records, task completion and other intervention data for compliance measure/IPE	University of Hertfordshire
May-July 2021	Interviews in case study schools (teachers, subject leaders, headteachers, governors) and documentary review	RAND Europe
June – July 2021	Survey of headteachers all schools	RAND Europe
June – July 2021	Outcome testing (Cohort B) and collection of surveys from Year 5 teacher and subject leaders	NFER/RAND Europe
15 February 2022	Draft EEF report	RAND Europe
September 2022	Final EEF report	RAND Europe

***These milestones were revised following trial adaptations due to COVID-19**

References

Allison, P. D., (2012) *Handling Missing Data by Maximum Likelihood.* Haverford, PA: Statistical Horizons. Retrieved from: https://statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf

Blouin, D., Teikan, A., Kamin, C., Harris, & Harris, I. B. The impact of accreditation on medical schools' processes. *Medical Education,* 52(2), 182-191. doi: 10.1111/medu.13461

Blumenfeld, P., Modell, J., Bartko, W. T., Secada, W., Fredricks, J., Friedel, J., et al. (2005). School engagement of inner city students during middle childhood. In C. R. Cooper, C. Garcia Coll, W. T. Bartko, H. M. Davis, & C. Chatman (Eds.), Developmental pathways through middle childhood: Rethinking diversity and contexts as resources (pp. 145–170). Mahwah, NJ: Lawrence Erlbaum

Boyle, A., Taylor, A., Giacomantonio, C. & Sutherland, A. (2015) Using ambulance data to reduce community violence: critical literature review. *European Journal of Emergency Medicine,* 23(4), 248-252.

CFE (2017). 'State of the nation' report of UK primary science education. Baseline research for the Wellcome Trust Primary Science Campaign. https://dera.ioe.ac.uk/31511/1/state-of-the-nation-report-of-uk-science-education-1.pdf

CFE (2019). Understanding the 'state of the nation' report of UK primary science education. A baseline report for the Wellcome Trust. https://wellcome.ac.uk/sites/default/files/understanding-state-of-the-nation-report-of-uk-primary-science-education.pdf

Cundill, B., & Alexander, N. D. (2015). Sample size calculations for skewed distributions. BMC medical research methodology, 15(1),28. DOI 10.1186/s12874-015-0023-0) .

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017) *Effective Teacher Professional Development.* Palo Alto, CA: Learning Policy Institute. Retrieved from: https://learningpolicyinstitute.org/product/teacher-prof/dev

Dong, N., & Maynard, R. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness,* 6(1), 24-67. doi: 10.1080/19345747.2012.673143

Dong, N., Kelcey, B., Spybrook, J., & Maynard, R. A. (2017). PowerUp!-Moderator: A tool for calculating statistical power and minimum detectable effect size of the moderator effects in cluster randomized trials (Version 1.08) [Software]. Available from http://www.causalevaluation.org/

Education Endowment Foundation (2015)*EEF Guidance on Cost Evaluations.* London: Education Endowment Foundation. Retrieved from: https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_guidance_to_evaluators_on_cost_evaluation.pdf

Education Endowment Foundation (2018) *Statistical Guidance for EEF evaluations.* London: Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In Handbook of research on student engagement (pp. 763-782). Springer, Boston, MA.

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., Garrett, R., Yang, R., Borman, G. D., & Wel, T. E. (2016). Focusing on mathematical knowledge: The impact of content-intensive teacher professional development (NCEE 2016-4010). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., Doolittle, F., Warner, E., (2011). Middle School Mathematics Professional Development Impact Study: Findings After the Second Year of Implementation (NCEE 2011-4024). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017) Effects of professional development on the quality of teaching: Results from a randomised controlled trial of Quality Teaching Rounds, *Teaching and Teacher Education,* 68, 99-113. doi: 10.1016/j.tate.2017.08.007

Hanbury, N., Prosser, M. & Rickinson, M. (2008) The differential impact of UK accredited teaching development programmes on academics' approaches to teaching. *Studies in Higher Education,* 33(4), 469-483, doi: 10.1080/03075070802211844

Hanley, P., Slavin, R., & Elliott, L. (2015). Thinking, Doing, Talking Science: Evaluation Report and Executive Summary. *Education Endowment Foundation*.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials, *British Medical Journal,* 343(d5928)

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K., (2016) *Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook.* Education Endowment Foundation.

Kind, P., Jones, K., & Barmby, P. (2007). Developing attitudes towards science measures. *International Journal of Science Education 29*(7), 871–893.

Kitmitto, S, González, R.,Mezzanote, J. & Che, Y. (2018) Thinking, Doing, Talking Science: Evaluation report and executive summary. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/TDTS.pdf

Oakes, J. M. (2013) Effect identification in comparative effectiveness research. *The Journal for Electronic Health Data and Methods,* 1(1):1004. doi: 10.13063/2327-9214.1004

Ofsted (2013) *Maintaining curiosity: A survey into science education in schools.* Ofsted. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/379164/Maintaining_20curiosity_20a_20survey_20into_20science_20education_20in_20schools.pdf

Ofsted (2019). Intention and substance: further findings on primary school science from phase 3 of Ofsted's curriculum research. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/777992/Intention_and_substance_findings_paper_on_primary_school_science_110219.pdf

Opfer, D. (2016). Conditions and practices associated with teacher professional development and its impact on instruction in TALIS 2013. OECD Education Working Paper, No. 138. OECD Publishing

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). Educational and Psychological Measurement, 53 (3), 801–813.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. The Journal of Educational Research, 95(6), 323-332.

Sims, S. and Fletcher-Wood, H. (2018) Characteristics of effective teacher professional development: what we know, what we don't, how we can find out. UCL Institute of Education. Available at: https://improvingteaching.co.uk/wp-content/uploads/2018/09/Characteristics-of-Effective-Teacher-Professional-Development.pdf

Skinner, E. A., Marchand, G., Furrer, C., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic. Journal of Educational Psychology, 100(4), 765–781. doi:10.1037/a0012840.

Speckesser, S., Runge, J., Foliano, F., Bursnall, M., Hudson-Sharp, N., Rolfe, H. & Anders, J. (2018) *Embedding Formative Assessment: Evaluation report and executive summary.* Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/public/files/EFA_evaluation_report.pdf

Spybrook, J., Kelcey, B., & Dong, N. (2016). Power for detecting treatment by moderator effects in two and three-level cluster randomized trials. *Journal of Educational and Behavioral Statistics.* doi: 10.3102/1076998616655442

Sutherland, A., Strang, L., Stepanek, M., Giocmantonio, C., & Boyle, A., (2017) *Using ambulance data for violence prevention: technical report.* Santa Monica, CA: RAND Corporation

Volkwein, J. F., Rattuca, L. R., Harper, B. J., & Domingo, R. J. (2006), Measuring the impact of professional accreditation on student experiences and learning outcomes. *Research in Higher Education,* 48(2), 251-282. doi: 10.1007/s11162-006-9039-y

White, E., Dickerson, C., & Mackintosh, J. (2015) *Impact of Royal Society of Chemistry bursary-funded Primary Science Quality Mark on primary science teaching: final report.* University of Hertfordshire. Retrieved from: https://www.researchgate.net/publication/305033454_Evaluation_of_the_Primary_Science_Quality_Mark_programme_2013-15.

White, Elizabeth & Dickerson, Claire & Mackintosh, Julia & Levy, Roger. (2016). Evaluation of the Primary Science Quality Mark programme 2013-15. DOI: 10.13140/RG.2.1.1312.9209https://www.researchgate.net/publication/305033454_Evaluation_of_the_Primary_Science_Quality_Mark_programme_2013-15

White, E., Dickerson, C., Mackintosh, J., & Levy, R. (2016) *Evaluation of the Primary Science Quality Mark programme – 2013-15.* University of Hertfordshire. Retrieved from: http://researchprofiles.herts.ac.uk/portal/en/projects/evaluation-of-rsc-bursary-funded-psqm(cdd1fbd0-0609-48f8-a554-c6f910a543c8).html

Wellcome Trust (2014). Primary Science: Is It Missing Out? Recommendations for reviving primary science. Retrieved from: https://wellcome.ac.uk/sites/default/files/primary-science-is-it-missing-out-wellcome-sep14.pdf