Trial Evaluation Protocol
Leadership Lite
Evaluator (institution): National Foundation for Educational Research (NFER)
Principal investigator(s): Jack Worth

Education
Endowment
Foundation

PROJECT TITLE	**Evaluation of Leadership Lite**
DEVELOPER (INSTITUTION)	Carmel Education Trust (CET)
EVALUATOR (INSTITUTION)	National Foundation for Educational Research (NFER)
PRINCIPAL INVESTIGATOR(S)	Jack Worth
PROTOCOL AUTHOR(S)	Palak Roy
TRIAL DESIGN	Two-arm cluster randomised controlled trial with random allocation at school level
TARGET GROUP	Science teachers in participating secondary schools
NUMBER OF SCHOOLS	Up to 140 secondary schools across two cohorts
NUMBER OF TEACHERS	8 Science teachers in each school
PRIMARY OUTCOME	Science teacher retention in state-funded schools
SECONDARY OUTCOME	Science teacher retention in the same school as at randomisation Student GCSE attainment in science Teacher workload and job satisfaction Student progression to science A-level (cohort 1 only)

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.0 [*original*]	17/05/2019	*[leave blank for the original version]*

Table of contents

Intervention

Brief name

Leadership Lite.

Why (rationale/theory)

Leadership Lite is a leadership development and school improvement intervention aiming to reduce teacher workload and increase teacher satisfaction and, ultimately, teacher retention in the profession. Improving student attainment outcomes is also a goal of the school improvement programme. Leadership Lite was developed and piloted by Carmel Education Trust – a Multi-Academy Trust and Teaching School in the North East of England. The programme, delivered over two years, focuses on three aspects of the quality of provision in science departments: quality assurance procedures; marking and feedback; and classroom practice. The intervention uses a series of evidence-based school improvement strategies that have workload reduction as a core principle.

The programme supports senior and middle leaders (heads of science), as well as governors to develop the skills of effective leadership and culture change management (e.g. Goleman 2000; Hill *et al.*, 2016a and b) to support the implementation of evidence-based school improvement strategies within science departments. Science teachers also receive instruction on implementing the strategies within their own practice. The strategies and approaches advocated in the Leadership Lite programme draw on substantial evidence of effective practices in relation to:

- quality assurance procedures – emphasising informal, supportive peer-observation and lesson demonstrations, and greater prominence of student voice to understand the effectiveness of lessons, based on evidence from Coe et al., 2014; Timperley et al., 2007; and Joyce and Showers, 2002.

- marking and feedback – reducing the quantity of written teacher marking and increasing the focus on student peer- and self-assessment, and the impact of marking on students' understanding of how they can improve, based on evidence from Shute, 2008; Hattie and Timperley, 2007; and Black and Wiliam, 1998.

- classroom practice – emphasising teaching focused on clear goals, responsiveness to students' prior knowledge and misconceptions, learning activities that deepen thinking and facilitate metacognition, as well as reducing the unnecessary detail of lesson planning, and shifting from individual to more reflective and shared lesson planning, based on evidence from Black and Wiliam, 1998; Sadler, 1989; Driver et al., 1994; Hattie, 2009; Quigley et al., 2018; and Adey and Shayer, 1994.

Who (recipients)

In each participating school the programme involves: a senior leader, a governor, the head of the science department, a nominated 'lead teacher' from the science department who will support the implementation of the approaches, as well as all other science teaching staff, including science subject leads, science teachers and science teachers who are newly and recently qualified (NQTs/RQTs) and early career teachers (ECTs) in their first five years of teaching.

What (materials)

Core programme components include: direct face-to-face training; gap tasks in between training sessions; network meetings; additional in-school support as required; exemplar materials, policies and templates (e.g. sample lesson plans); and online support (e.g. web-based moodle). Exemplar materials, policies and templates are designed to provide teachers and leaders with 'ready-to-use' resources to support the implementation of the strategies advocated in the training with minimal additional input from participants.

What (procedures/activities/processes)

The training for senior and middle leaders and governors (as well as the nominated 'lead teacher', if appropriate) involves practical activities designed to encourage reflection on leadership styles and culture, and promote evidence-based leadership approaches (e.g. understanding how to influence individuals with different motivations; identifying effective leadership attributes). This is designed to enable these participants to lead the improvements in the science department around quality assurance practices, marking and feedback, and classroom practice in order to reduce workload and increase effectiveness.

The training for science teachers focuses primarily on developing classroom practice (e.g. effective strategies for lesson planning), though it also touches on the other two aspects of the programme - quality assurance, and marking and feedback – and the implications of the Leadership Lite advocated practices for classroom science teaching (e.g. understanding of student peer- and self-assessment techniques, strategies to support moderation, and how to maximise the benefit of peer coaching).

Gap tasks are assigned in between the training sessions to encourage participants to implement the ideas and approaches presented at the training, and to feed back about their experiences at subsequent sessions and via the moodle (e.g. a brief reflection piece having tried a recommended approach, watching a video tutorial of a lesson and writing a short reflection having modelled the lesson). Further in-school support is also available for schools that require additional support with implementation and organisational changes (e.g. in the form of diagnostics, coaching, advice, co-planning and demonstrations).

Network sessions provide an opportunity for participating schools to present and feedback on any changes they have made to practices, and facilitate school-to-school sharing and further embedding of the approaches.

Who (providers/implementers)

The Leadership Lite programme is delivered by a network of partners, mainly centred on Science Learning Partnerships, but may also include other school improvement organisations, such as Local Authorities and Teaching School Alliances. Science Learning Partnerships are led by local teaching school alliances, schools, colleges, higher education institutions and other local partners with expertise and excellence in science. Staff from Carmel Education Trust's Teaching School and Science Learning Partnership will train Specialist Leaders in Education (SLEs) and consultants from the network of Science Learning Partnerships to deliver the Leadership Lite programme to clusters of schools within their local areas. Each cluster will comprise between five to 16 schools in receipt of Leadership Lite. Each Leadership Lite school will be allocated a named SLE/consultant to

work with and each SLE/consultant will work with between one to six schools (hence, more than one SLE/consultant may work with each cluster of schools).

How (mode of delivery)

The Leadership Lite programme is delivered through:

- external face-to-face sessions for senior leaders and governors from all participating schools within each local area delivered at local training centres (i.e. off-site training delivered to 'clusters' of participating schools)

- external face-to-face training for middle leaders (Head of Science, science subject leads) from all participating schools within each local area delivered at local training centres (i.e. off-site training delivered to 'clusters' of participating schools)

- external face-to-face training for science teachers from all participating schools within each local area delivered at local training centres (i.e. off-site training delivered to 'clusters' of participating schools). In addition to this, in-school training for all science teachers may also be provided depending on the diagnostic.

- gap tasks in-between the training sessions to encourage implementation of the ideas and approaches presented at the training and pre-reading to maximise the face-to-face training opportunities. Brief summaries of participants' experiences of implementing the strategies will be uploaded to the programme moodle.

- local network sessions for senior leaders, middle leaders, governors, and classroom science teachers (face-to-face twilight meetings with option for remote dial-in)

- in-school support for implementation and organisational changes, as required (e.g. diagnostic, coaching, advice, co-planning and demonstration)

- course materials and exemplar policies and templates

- on-going online support will be provided to participants throughout the programme via a Virtual Learning Environment (moodle) (e.g. participants will be asked to upload completed gap tasks).

Where (location of intervention)

Training is delivered externally at local training centres and on-site in schools. Implementation of the strategies takes place in regular classrooms in participating schools. For the purposes of this trial, the Leadership Lite programme will be delivered to schools in the north of England and surrounding areas.

When and how much (duration and dosage)

The Leadership Lite programme is delivered over two years. Sessions for senior leaders and governors will take place once per term in the first year of the programme and will be for half a day per session (i.e. three x half day sessions). The expectation is that a minimum of one senior leader and one governor from each school attends each session. Training sessions for middle leaders (Heads of Science) and the 'lead teachers' will take place once per term in the first year of the programme and will be for a full day per session (i.e. three x full day training sessions). The expectation is that a minimum of one middle leader (Head of Science) or a designated 'lead teacher' per school attends each training session. Please note that the SLT/Governor sessions will take place on the afternoon of the Middle Leader

training sessions. Training sessions for science teachers and science NQTs/ECTs will take place once per term in the first year of the programme and will be for a full day per session (i.e. three x full day training sessions). The expectation is that at least one teacher attends each of the cluster training sessions. For the on-site training sessions that are scheduled (based on the diagnostic), the expectation is that as many science teaching staff as possible attend (this could also include the Head of Science and lead teacher if they wish to attend both the training for 'leaders' and for 'teachers'). In the second year of the programme, all participants will be able to join twilight network sessions which will be offered once per term (i.e. three x twilight network sessions). It is expected that a minimum of two participating members of staff from each school will attend each round of network sessions. Additional in-school support will be provided, in negotiation with each school, to support implementation of Leadership Lite approaches (e.g. diagnostic, coaching, advice, co-planning and demonstrations). In the second year of the programme, 'catch-up' training sessions will be offered for any new staff or staff who missed the training in the first year of the programme, as required.

There is some flexibility around who attends the programme training and networking events, depending on what is appropriate for each participating school. However, the tenets of the programme delivery are that it is an holistic approach facilitated by developing common approaches to school improvement that are understood by senior leaders and governors through to classroom teachers in order to support phased implementation and secure embedding of the approaches. Hence, it is expected that each school will involve one or more members of staff with each of the different activities offered.

Each participant should therefore experience a minimum of three training events and three network sessions, however some participants (such as middle leaders – Heads of Science, heads of biology/physics/chemistry, 'lead teachers') may receive more if they attend both the leaders' and teachers' training. In addition to attending training and networking events, each participant will be expected to engage with gap tasks and use the course materials. There are gap tasks between each training session (including leaders' and teachers' training) (i.e. two tasks per participant), as well as pre-course reading for all participants, and an end of training reporting task for middle leaders and teachers. Apart from the pre-reading, participation in the gap-tasks will be monitored by completion of brief reflective journals and other tasks that are uploaded to a programme moodle.

Tailoring (adaptation)

The Leadership Lite programme incorporates four dimensions that are considered as essential for effective school improvement and teacher satisfaction and retention: leadership of change; classroom practice; marking and feedback; and quality assurance. Within each dimension, Leadership Lite identifies a spectrum of practices ranging from those that are considered least effective through to more effective and evidence-based practices. These dimensions are central to the Leadership Lite programme and will underpin the delivery of the programme for all participants. However, there will be some variation in the emphasis placed on each dimension depending on the programme participants' roles and needs. Training for senior and middle leaders and governors will emphasise the 'leadership of change' aspects; conversely training for science teachers will emphasise the 'classroom practice aspects'. There will be scope for tailoring in terms of the specific examples of

practices that schools and participants elect to reform and the extent of this, depending on their individual needs. There will also be tailoring in the sense that individual schools will identify different priorities and areas of focus for the development and application of the practices.

Table 1 below sets out the core dimensions of Leadership Lite, the overarching focus of each dimension, and specific examples of practices and how Leadership Lite aims to shift these from common practices at one end of the spectrum to Leadership Lite practices at the other.

Table 1   Leadership Lite dimensions, strategies and example practices

Dimension	Overarching focus of each dimension	Specific example of common practice	Specific example of Leadership Lite practice
Leadership of change	Shifting from 'short-term leadership focused on improving performance' to 'strategic, people-focused long-term leadership focused on improving outcomes'	Leadership is based on the leader's vision	Leadership is based on a shared vision with a common goal (e.g. focused around the 'ideal learner')
		Leadership is focused on embedding established best practice across all areas of the school, regardless of the need, often leading to multiple initiatives.	Leadership is focused on essential change (e.g. based on diagnostic work)
		Imposing change that is felt necessary, such as in response to the current national agenda, perceptions of Ofsted etc.	Leading people using phased implementation, accounting for the strengths and weaknesses of the team and the enthusiasm for change of individuals/groups.
		Ensuring compliance via consequence as the main mechanism	Ensuring compliance via motivation as the main mechanism
		Improvements are focused on examination performance	Improvements are focused on the broader outcomes for children that include skill development, employability etc.
		Leaders keep up the appearance of the school	Leaders serve the community
		Teachers are seen as part of the problem	Teachers are seen as part of the solution
		Short-term wins are prioritised, sometimes at the expense of long-term gains, e.g. teachers not released for CPD as they are needed in the classroom, intervention/extra classes for Year 11 take place at the expense of earlier intervention	Long-term approaches are prioritised, such as investment in CPD, intervention across all year groups
		Individual accountability	'Collaborative' accountability

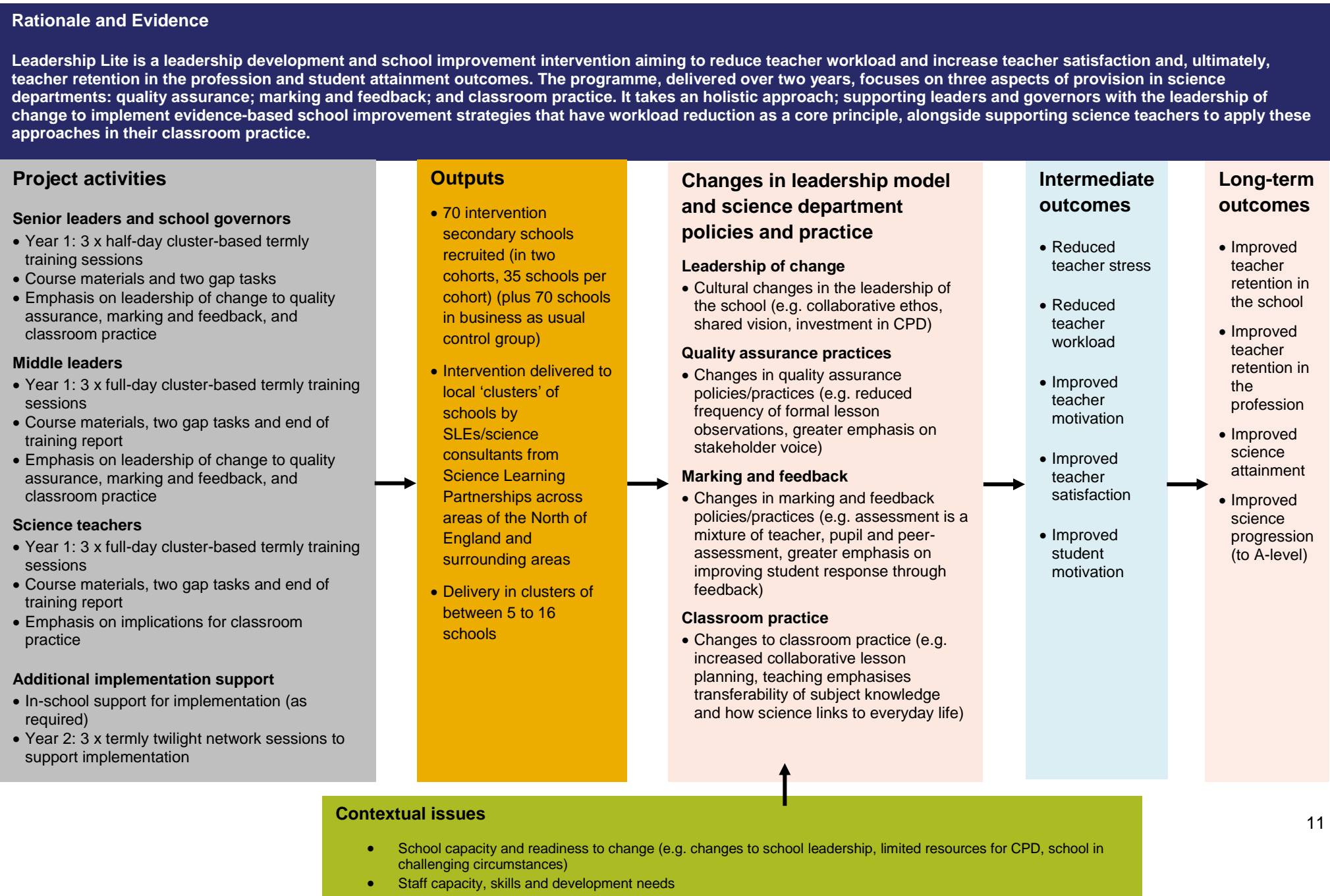| Classroom practice	Shifting from 'classroom practice is focused on the teachers planning their lessons well to teach the students what they need to know in order to perform well in examinations' to 'classroom practice is focused on collaborative and essential planning, designed to develop students skills and the transferability of those to develop an enthusiasm for science and an understanding of how science fits in to the world and careers'	Lone planning	Collaborative planning
		Detailed lesson planning	Minimalistic planning focuses on the necessary
		Teaching is focused on developing subject knowledge and understanding	Teaching is focused on developing the transferability of the subject knowledge, skills and understanding
		Teaching is focused on covering the curriculum	The curriculum is designed to widen the understanding of how science fits in to the wider world, everyday life and future careers.
		Ensuring student compliance via consequence	Ensuring student compliance through motivation, engagement and effective practice
		Teaching is task focused. Tasks are adapted for the ability of the students	Teachers act professionally and adapt their work to fit the needs of the students, without necessarily writing or recording this
		Tasks are clearly differentiated according to ability	Activities are both accessible and challenging for all students
Marking and feedback	Shifting from 'assessment, marking and feedback is the process for teachers to diagnose and guide students to improve their work' to 'assessment, marking and feedback encourages high levels of student response, better quality work and independence'	Assessment and feedback is focused on detailed teacher marking	Assessment and feedback is a process to diagnose and improve students' work and can happen interactively, in-class and may not have a written record
		Data is trusted implicitly and fine differences in the data are used to make decisions. Few processes (if any) are in place to increase the quality of the data. Quantity of data may be prioritised over quality	There is an understanding of the tentative nature of data and leaders respond accordingly to make it more robust, for example, by moderation (internal and external), standardisation across assessments, fewer but higher quality assessments, referencing with a larger sample size
		There is a focus on regular (e.g. fortnightly) detailed diagnostic teacher comments as the main method of feedback	Assessment, marking and feedback is a mixture of teacher, peer and self-assessment. These skills are taught to children with the aim of them becoming self-remediating/self-regulating. Check lists/highlighted success criteria and/or symbolic marking may also feature
		There is a school/departmental focus on the quality and quantity of teacher feedback. This	There is a focus on improving the student response through assessment, marking and feedback

		might include meeting established best practice criteria	
		Feedback is usually given in the form of 'what went well' then even better if, before students are expected to respond	Assessment, marking and feedback focuses on improvement first (e.g. uses the 'even better if (ebi)' then at some later point there may be 'what went well' (www) or 'mission then medal'
Quality assurance	Shifting from 'quality assurance is a process to check on the work of teachers. It typically centres on lesson observation (both formal and informal) and other 'input' measures' to 'quality assurance procedures are designed not to interfere with the work of teachers, and to triangulate evidence from outcome data and stakeholder perceptions. Lesson observations (formal and informal) are rarely, if ever, used as part of quality assurance processes. Instead they are carried out by teachers who observe their peers in a developmental way'	Broad or generic interventions based on established best-practice	Focused interventions based on themed question level analysis (TQLA)
		Focus on features of practice meeting particular 'ideal' criteria (for example, lesson planning, homework, marking, may have checklists of particular features, upon which QA is based)	Focused on the impact on outcomes of provision in school (e.g. lessons, home learning, marking/feedback) rather than checking that these have particular features
		Little use of stakeholder voice	Stakeholder voice is used regularly (e.g. termly) as part of outcome measures (e.g. students, staff, parents)
		Implied distrust of a significant number of teachers	Implied trust in the professionalism of the vast majority of teachers
		Book scrutiny takes place by asking for random sample of books (or sample of specific students)	Book review takes place as part of a dialogue with students (i.e. students are present during the review and are asked questions to probe assumptions about practice related to marking and feedback)
		Superficial use of collected data, which is often collected more than three times per year	Deeper and smarter use of data, which has been through a process to ensure that it is reliable
		Intensive monitoring with multiple formal lesson observations (typically termly or more frequently) and informal observations (such as learning walks), which may take place as frequently as fortnightly or even weekly	Light touch monitoring, with fewer formal lesson observations (ideally zero or close to zero) and fewer informal lesson observations (ideally zero or close to zero). Teachers may choose to demonstrate their practice to middle/senior leaders via demonstration lessons

The programme aims to maximise effective implementation by providing initial central training for the delivery team from Carmel Education Trust staff who have extensive experience of teacher development and school improvement. The delivery team is comprised of experienced science teachers and leaders, with significant experience of leading professional development and providing advisory support. Trainers are expected to hold some form of accreditation in science teaching, leadership or CPD delivery (e.g. STEM Learning Lead, ASE Chartered Science teacher). The delivery team will be provided with a package of presentation and course materials that have been developed centrally by the Carmel Education Trust team and will be used consistently by the delivery team throughout the programme. Carmel Education Trust will monitor and Quality Assure the programme delivery to ensure the Leadership Lite training is delivered as intended. Participating schools will be provided with clear information and guidance about the programme and the strategies and practices to implement. The programme has been refined based on learning from a pilot study of the implementation of Leadership Lite in maths departments in a small number of schools. Implementation issues arising in the pilot study (e.g. low levels of engagement with particular aspects of the support, schools' reluctance to reduce the number of lesson observations) have been mitigated through modifications to the intervention design and materials.

Figure 1 below sets out a logic model for the evaluation of Leadership Lite, outlining the activities and outputs of the programme, and the changes to practice and intermediate outcomes that are expected to lead to ultimate outcomes on teacher retention and student attainment and progression.

Figure 1: Evaluation Logic Model – Leadership Lite

Rationale and Evidence

Leadership Lite is a leadership development and school improvement intervention aiming to reduce teacher workload and increase teacher satisfaction and, ultimately, teacher retention in the profession and student attainment outcomes. The programme, delivered over two years, focuses on three aspects of provision in science departments: quality assurance; marking and feedback; and classroom practice. It takes an holistic approach; supporting leaders and governors with the leadership of change to implement evidence-based school improvement strategies that have workload reduction as a core principle, alongside supporting science teachers to apply these approaches in their classroom practice.

Project activities

Senior leaders and school governors
- Year 1: 3 x half-day cluster-based termly training sessions
- Course materials and two gap tasks
- Emphasis on leadership of change to quality assurance, marking and feedback, and classroom practice

Middle leaders
- Year 1: 3 x full-day cluster-based termly training sessions
- Course materials, two gap tasks and end of training report
- Emphasis on leadership of change to quality assurance, marking and feedback, and classroom practice

Science teachers
- Year 1: 3 x full-day cluster-based termly training sessions
- Course materials, two gap tasks and end of training report
- Emphasis on implications for classroom practice

Additional implementation support
- In-school support for implementation (as required)
- Year 2: 3 x termly twilight network sessions to support implementation

Outputs
- 70 intervention secondary schools recruited (in two cohorts, 35 schools per cohort) (plus 70 schools in business as usual control group)
- Intervention delivered to local 'clusters' of schools by SLEs/science consultants from Science Learning Partnerships across areas of the North of England and surrounding areas
- Delivery in clusters of between 5 to 16 schools

Changes in leadership model and science department policies and practice

Leadership of change
- Cultural changes in the leadership of the school (e.g. collaborative ethos, shared vision, investment in CPD)

Quality assurance practices
- Changes in quality assurance policies/practices (e.g. reduced frequency of formal lesson observations, greater emphasis on stakeholder voice)

Marking and feedback
- Changes in marking and feedback policies/practices (e.g. assessment is a mixture of teacher, pupil and peer-assessment, greater emphasis on improving student response through feedback)

Classroom practice
- Changes to classroom practice (e.g. increased collaborative lesson planning, teaching emphasises transferability of subject knowledge and how science links to everyday life)

Intermediate outcomes
- Reduced teacher stress
- Reduced teacher workload
- Improved teacher motivation
- Improved teacher satisfaction
- Improved student motivation

Long-term outcomes
- Improved teacher retention in the school
- Improved teacher retention in the profession
- Improved science attainment
- Improved science progression (to A-level)

Contextual issues
- School capacity and readiness to change (e.g. changes to school leadership, limited resources for CPD, school in challenging circumstances)
- Staff capacity, skills and development needs

11

Study rationale and background

Ensuring sufficient supply of teachers in secondary schools is an issue of increasing policy importance due to the number of pupils forecast to increase by 15 per cent between 2018 and 2025. Long-standing under-recruitment of science teachers compared to the numbers needed to maintain sufficient supply, means that the supply of science teachers is of even greater relevance. Recent research has found that science teachers are more likely to leave the profession than non-science teachers, particularly within their first five years of teaching (Worth and De Lazzari, 2017; DfE, 2017). While several government schemes have sought to improve the recruitment and retention picture, largely through offering financial incentives to physics and chemistry teachers, but also with CPD and support packages, they have had little success (Public Accounts Committee, 2018). There is a dearth of causal evidence as to how to improve teacher retention in England.

Leadership Lite is a school-developed approach that has been piloted in a small number of schools (Bailey, forthcoming). The combined approach thus does not have a strong evidence base, but is rooted in a range of evidence-backed strategies as well as the team's practical understanding of school improvement. These strategies are clearly defined and based on theories that are popular with schools. Therefore, a randomised controlled trial is required to evaluate whether the programme can improve science teacher retention when implemented at a large scale.

Impact Evaluation

Research questions

Primary research question for this trial is:

1. What is the impact of Leadership Lite on retention of science teachers in the state-funded schools in England six months after the end of programme delivery?

Secondary research questions for this trial are:

2. What is the impact of Leadership Lite on retention of teachers in the school that they were employed in at randomisation within six months of the end of programme delivery?

3. What is the impact of Leadership Lite on students' GCSE attainment in science?

4. What is the impact of Leadership Lite on teachers' workload and job satisfaction?

5. What is the impact of Leadership Lite on student progression to science A-level (Cohort 1 only)?

Trial type and number of arms		Two-arm, cluster randomised
Unit of randomisation		School
Stratification variables (if applicable)		Up to five geographic areas and randomisation cohort
Primary outcome	variable	Science teacher retention in state-funded schools
	measure (instrument, scale)	Teacher records from School Workforce Census (binary indicator: 1=longitudinal record appears in SWC database, 0=does not appear).
Secondary outcome(s)	variable(s)	Science teacher retention in the same school as at randomisation
	measure(s) (instrument, scale)	Teacher records from School Workforce Census (binary indicator: 1=longitudinal record in the same school in SWC database, 0=has moved school or does not appear in SWC).
Secondary outcome(s)	variable(s)	Student GCSE attainment in science
	measure(s) (instrument, scale)	National Pupil Database (GCSE point score)
Secondary outcome(s)	variable(s)	Student A-level progression in science
	measure(s) (instrument, scale)	National Pupil Database (binary indicator: 1=entry to an A-level in a science subject, 0= no science entry)
Secondary outcome(s)	variable(s)	Teacher workload and job satisfaction
	measure(s) (instrument, scale)	Teacher pre- and post-survey. Scales will be defined using the baseline survey data.

As Leadership Lite is a whole-school intervention with a department focus, the trial will be a two-armed randomised controlled trial with randomisation (to control and intervention) at school level. This will eliminate the challenges of implementing a teacher-level design and the risk of contamination. It will be a stratified randomisation of schools to intervention and control. The stratification is planned to achieve geographic balance, and thereby support effective intervention delivery. The recruitment and randomisation will be in two cohorts to support intervention delivery.

We will access the primary outcome measure from the School Workforce Census (SWC). At baseline, we will collect identifying information about all science teachers from participating schools. This will include names, dates of birth, teacher reference numbers (TRN), role in the science department of all their science teachers, as a condition of randomisation. This list will be matched to the SWC to gather data on teachers' characteristics, and enable retention outcomes to be followed up longitudinally in subsequent censuses. Once the schools submit the administrative teacher data, NFER will reveal the school's randomisation outcome to them.

A first cohort of up to 70 schools will be randomly allocated to either intervention or control in May-June 2019 and a second cohort of up to 70 schools will be randomised in May-June 2021 (see 'participants' section below for more details). Schools assigned to the intervention group will receive the Leadership Lite programme for two years. For each participating school the programme will involve:

- Senior leaders and governors: termly sessions in the first year (three half day sessions) and termly twilight network sessions in the second year.

- Middle leaders/Heads of Science: termly training sessions in the first year (three full day training sessions) and termly twilight network sessions in the second year.

- Science teachers/NQTs: termly training sessions in the first year (three full day training sessions) and termly twilight network sessions in the second year.

Schools in the control group will continue with their usual practice and will receive a payment of £1500 in lieu of the intervention, following completion of data collection activities.

Randomisation

Once schools submit administrative teacher data, NFER will reveal the school's randomisation outcome to them. An NFER statistician will randomise schools from the first cohort of up to 70 schools in May-June 2019 and from the second cohort of up to 70 schools in May-June 2021. The statistician will not be blinded to group allocation. Randomisation will be stratified by two or three (maximum five) broad geographical areas to aid effective intervention delivery. Within each stratified group, schools will be allocated half to the intervention group and half to the control group. Schools that are part of the same school structure (e.g. a multi-academy trust) will be randomised to the same group as a single unit.

Participants

Schools
CET will be responsible for school recruitment. They will aim to recruit 140 secondary schools across two cohorts: 70 schools in cohort 1 and 70 schools in Cohort 2. All secondary schools in the North of England and surrounding areas will be eligible to take part in the trial. School recruitment for cohort 1 will be January-April 2019 and for Cohort 2 will be November 2020-March 2021. Schools will sign up to the trial via a memorandum of understanding (MOU) where they will nominate a key contact person. Schools will be recruited via a variety of means, including emails from our trust, delivery and other partners, promotion at local events for science and leaders and direct contact. Once a school has expressed an interest, steps will be taken in the communication at each stage to ensure that they understand the nature of the trial.

Teachers
Once a school signs the MOU, we will collect science teachers' identifying information in order to match to the SWC and access the primary outcome measure in future censuses. It will be a condition of randomisation for schools to provide the teacher information. We will aim to include all science teachers who are employed in the academic year 2019/20 (2021/22 for Cohort 2). We will include those on maternity leave but exclude staff who are known to be leaving during the 2018/19 academic year.

Science teachers will include those who are teaching Physics, Chemistry, Biology or General Science to any year group for more than 50 per cent of their timetable and for a minimum of four hours per week for part-time staff in the academic year 2019/20 (2021/22 for Cohort 2). In case there are staffing changes, schools will send us an update by September 2019 (prior to intervention delivery). We will not measure the retention outcomes of science teachers who join the school after the intervention delivery starts for a given cohort: this eliminates the risk of their recruitment being for a reason that is biased by the intervention.

Pupils

We will gather GCSE science attainment data from the NPD as a secondary outcome measure. This will include students who are in Year 9 at randomisation (academic year 2018/19 for Cohort 1 schools and 2020/21 for Cohort 2 schools), and therefore sitting GCSEs at the end of the two-year intervention (May 2021 for Cohort 1 and May 2023 for Cohort 2).

Sample size calculations

Table 2:   Sample size calculations

		OVERALL
MDES		5.8 percentage points (MDES = 0.2) (Odds ratio = 1.4) (Control group retention rate assumed = 77%)
Covariate correlations	level 1 (teacher)	0.32
	level 2 (school)	0
Intra-cluster correlations (ICCs)	level 2 (school)	0.08
Alpha		0.05
Power		0.8
One-sided or two-sided?		Two-sided
Average cluster size		8 teachers per school
Number of schools	Intervention	70
	Control	70
	Total	140
Number of teachers	Intervention	560
	Control	560
	Total	1120

Figure 2 shows the sample size calculations, assuming 140 schools are recruited in total. Using parameters estimated from teacher-level SWC data, we estimate that an evenly-randomised sample of 140 schools will enable us to detect a difference of 5.8 percentage points with 80 per cent power (purple curve). Assuming a 77 per cent teacher retention rate

in the control group, the design gives adequate statistical power to detect an increase in the teacher retention rate to 83 per cent (or higher), or a reduction in the teacher retention rate to 71 per cent.

If recruitment of 140 schools is not achieved, then a design with fewer schools would still be feasible. For example, a design including 110 schools (55 intervention and 55 control) would result in a fairly small loss of power compared to a design with 140 schools. A design with 110 schools would enable us to detect a difference of 6.4 percentage points with 80 per cent power (blue curve).

The Leadership Lite documents suggested that the intervention was associated with a 23 percentage point increase in school retention rates in a pilot study Bailey, forthcoming). However, quasi-experimental research by Allen and Sims (2017) found that the impact of participating in a science CPD programme was associated with an increase of 3-4 percentage points in department-level rates of science teacher retention in the profession. This suggests that lowering the MDES from 5.8 to 6.4 percentage points would appreciably increase the risk of a false negative.

Figure 2   Statistical power curves

Power - n(intervention)=70; n(control)=70
Power - n(intervention)=55; n(control)=55
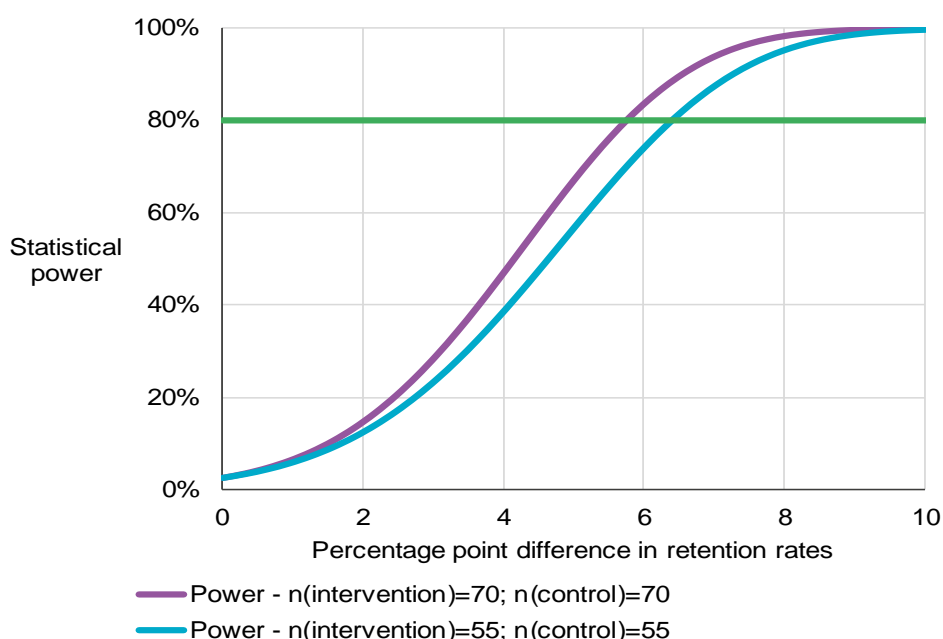
Assumptions:
8 science teachers per school (from analysis of SWC data)
ICC = 0.08 (from SWC)
Correlation with covariates (5-year age bands) = 0.32 (from SWC)
Confidence level of the test = 95 per cent
3-year science teacher retention rate for control schools = 77% (from SWC)

Outcome measures

Primary outcome

Teacher retention will be the main outcome measure, but it can be defined in a number of different ways. The main distinction is between retention in the profession and retention in the school. Both are important, but for different stakeholders: the number of teachers leaving

the profession affects the overall supply of teachers and therefore government policy, whereas staff turnover (teachers leaving the profession or moving school) affects schools' staffing and costs of recruitment. Because of the current concern over retaining enough science teachers in the profession, the primary outcome will be retention of teachers in the state-funded teaching workforce, and retention in the school will be included as a secondary outcome.

The primary outcome for this trial is the retention of science teachers in the state-funded school system six months after the end of programme delivery. This will be measured using data from the SWC. It will be a dichotomous (yes/no) variable, denoting each teacher's presence in the SWC, whether in the same school or in a different school, and teaching any subject[1]. The SWC is collected each November, so we can track the retention rate of the intervention and control cohorts of teachers over time. We will select retention in one particular post-intervention year as the primary outcome. As Leadership Lite is a two-year programme, the primary outcome will be in the census collected after the programme finishes: i.e. November 2021 for Cohort 1 and November 2023 for Cohort 2. Retention rate six months after the end of programme delivery is chosen as the primary outcome so that the impact of the intervention can be ascertained in a timely manner without waiting excessively for subsequent censuses. We will conduct exploratory analysis of the impact on subsequent post-intervention retention rates for Cohort 1: November 2022 SWC (18 months after end of delivery) and November 2023 SWC (30 months after end of delivery).

Secondary outcomes

There are a number of secondary outcome measures for this trial. The first one is the retention of science teachers in the same school that they were employed in at randomisation, six months after the end of programme delivery for each cohort. This will be measured via the same SWC censuses as the primary outcome measure, i.e. November 2021 and November 2023. It will be a dichotomous outcome variable, denoting each teacher's presence in the same school teaching any subject. In addition to this, we will also check whether the science teachers at the start of the trial were still teaching science six months after the end of programme delivery. To do so, we will use SWC data on which subject a teacher teaches as a sensitivity analysis, although the subject data only covers around two-thirds of science teachers because of missing SWC data. Therefore, the analysis will only include those for whom we have the subject data.

Another secondary outcome measure is science attainment - the impact of Leadership Lite programme on students' GCSE attainment in science. As the programme is a school-improvement programme, it is also important to consider its effects on students. The data will be accessed via the National Pupil Database (NPD)and include the entire Year 9 cohort at the time when the school was randomised- May/June 2019 for Cohort 1 and May/June 2021 for Cohort 2. As all students in a school will be given the same treatment, we do not need to collect any personal data for students. Therefore, the analysis will use de-identified NPD for all randomised schools. This attainment outcome will be an immediate impact of the two-year programme as the attainment data relates to the same academic year when the programme delivery finishes. We will use a GCSE Science point score measure, capturing

---

1. The primary outcome measure will be based only on SWC contract data, so will suffer from virtually no measurement attrition. We believe that this is an appropriate measure given the concern is about retaining teachers in the state-funded school system than teaching a different subject.

science attainment in a comparable way depending on which science GCSEs students sit (e.g. double or triple science). The current NPD data tables do not indicate the level of detail for these variables under the reformed qualifications: the measures will be described fully in the Statistical Analysis Plan (SAP).

Progression to A-level will also be a secondary outcome. As the only way to measure progression to studying a subject at A-level form administrative data is from the A-level outcome data, we will only be able to analyse this for cohort 1. Progression will be measured as a dichotomous variable (yes/no), measuring whether or not each student enters A-level examinations in a science subject.

We will also measure teachers' workload and job satisfaction as secondary outcome measures. These will be measured using the teacher surveys, which will be administered at baseline (before the intervention delivery) and follow-up (summer 2021 for cohort 1; summer 2023 for cohort 2). Where we combine multiple items to form scales that are used for outcome measurement, we will define these using factor analysis of the baseline survey data, as the survey items do not come from an existing instrument. We will class factors as reliable if they have a cronbach's alpha of 0.7 or above. If we do not find any reliable factors then we will simply produce frequencies for each group (intervention vs. control). By defining the scale on data collected before the intervention delivery, we ensure that there is no bias in the way it is defined. The scales will be defined fully in the SAP.

Analysis plan

Teacher retention analysis

The primary analysis will be intention-to-treat. We will analyse teacher data using either a multilevel logistic regression model or a survival analysis approach (e.g. Cox regression model). Our choice will be fully described and justified in the SAP. The school level in the model will account for clustering of teachers within schools. To maximise the degrees of freedom we have for our analysis we will minimise the number of teacher-level covariates in the model by only including age and experience as explanatory variables. NFER's previous research has shown that these variables explain most of the explainable variation in teacher retention rates. Randomised group, geographical stratifier and the recruitment cohort will be included as school-level covariates. We will follow the same structure of the model to analyse teacher retention in schools (secondary outcome).

Analysis of student outcomes

Student-level GCSE attainment analysis will also be intention-to-treat. This will be a multilevel regression model with schools and students as two levels. GCSE science attainment (point score) will be regressed on a number of covariates that are standard to the EEF approach to trial analysis: randomisation group, geographical stratifier, the recruitment cohort and pupil Key Stage 2 attainment as prior attainment. We will also analyse progression to science A-level, for cohort 1 only using a multilevel logistic regression model. Similar to the student attainment analysis, the outcome measure will be regressed on randomisation group, geographical stratifier, the recruitment cohort and pupil Key Stage 2 attainment as prior attainment.

Additional analyses

We will conduct on-treatment and/or instrumental variable compliance analysis by combining information on schools' implementation of the intervention from the process evaluation. We will consider how the intervention impact varies between schools with different levels of measured intervention fidelity. Fidelity will be measured by the extent of engagement with the various programme activities including training sessions, network sessions and gap-tasks. The expected minimum compliance with the intervention in each school is participation in: training (i.e. at least one senior leader, one governor, the Head of Science or a designated 'lead teacher', and as many science teachers as possible each participate in three training sessions); network sessions (i.e. a minimum of two representatives from senior leadership/governance and middle leadership from each school attends three network sessions); and gap tasks (i.e. two tasks for senior leaders; three tasks for middle leaders and teachers). However, this analysis will be non-causal and is likely to be underpowered. The detail of our approach will be outlined in the SAP.

We will analyse the impact of the intervention on intermediate outcomes, such as teachers' workload and job satisfaction, measured using the teacher surveys. We will use a multilevel regression model with schools and teachers as two levels, and use measures from the baseline survey as covariates.

We will also conduct sub-group analysis to analyse the impact of the intervention on the retention rates of particular groups of interest. We will explore the impact of the intervention on early-career teachers: those in their first five years of teaching after qualifying, measured using teacher's date of first entry to teaching in the state sector in the SWC. The rationale for this sub-group analysis is that the intervention is targeted at addressing issues that particularly affect early-career teachers, such as effectively managing workload.

We will also explore the impact of the intervention on science teachers with a degree in a science subject, measured using teacher qualification data in the SWC and the Department for Education's (DfE) mapping of degree subjects to subject specialism. The rationale is that these teachers have a high value to the teacher workforce because of their subject knowledge. However, the SWC data on qualifications is not available for all teachers, so this analysis may have a smaller sample size than the main analysis because of missing data.

We will follow EEF's Statistical analysis guidance to calculate appropriate effect sizes for each analysis. When the outcome measure is a binary variable, we will present the impact estimate in a standardised form. We anticipate this will be reported as either an odds ratio, risk ratio or hazard ratio. This will depend on the exact analysis method, which will be defined in detail in the SAP. Where the outcome measure is a continuous variable (e.g., GCSE attainment), the numerator for the effect size calculation will be the coefficient of the intervention group from the multilevel model. The effect sizes will be calculated using the total variance without covariates, as the denominator i.e. equivalent to Hedges' g. Confidence intervals for each effect size will be derived by multiplying the standard error of the intervention group model coefficient by 1.96. These will be converted to effect size confidence intervals using the same formula as the effect size itself. We will include further details in the SAP.

Implementation and process evaluation

The process evaluation will complement the impact evaluation by providing insights as to how the implementation of the Leadership Lite programme may account for the impact findings. It will explore how the programme is delivered, the engagement of participants, the effectiveness of individual elements of the programme as well as the package as a whole, and identify implications for future application. Table 3 outlines the data that will be collected to address each of the dimensions recommended in the 'IPE for interventions in education settings handbook' (Humphrey *et al.*, 2016).

Table 3:   Data collection to address IPE dimensions

IPE dimension	RQ/focus of data collection	Source of data	Timing of data collection
Fidelity and adherence	• How is Leadership Lite intended to work?	Developer interviews	Developer interviews: Autumn 2019
	• To what extent does the delivery of Leadership Lite adhere to the intended approach? Is it delivered as planned? (e.g. cluster delivery model, external and in-school training, involvement of senior and middle leaders and teachers, attendance at training and networking events, engagement with course materials and gap tasks) • How consistent is delivery across the clusters and how closely does delivery adhere to the practices advocated by Leadership Lite?	MI data; developer interviews; observations; case studies	<u>MI Data</u>: termly C1 and C2 <u>Developer interviews</u>: Autumn 2019, Autumn 2021, and Summer 2023 <u>Observations</u>: C1 Autumn 2019; C2 Autumn 2021 <u>Case studies</u>: C1 Autumn 2019 and Summer 2021; C2 Autumn 2021 and Summer 2023
Dosage	• To what extent does each school engage with Leadership Lite, and within each school to what extent does each targeted participant participate in the different aspects of Leadership Lite? (e.g. how many and which staff in each school participate in training, networking, and gap tasks that are submitted to the programme Moodle?)	MI data	<u>MI Data</u>: termly C1 and C2
Quality of delivery	• What do school leaders and teachers think about the quality and effectiveness of delivery by Leadership Lite? (e.g. views on the programme structure, content, and quality of SLEs/consultants) • What are school leaders' and teachers' perceptions of the effectiveness of individual elements of the programme? (e.g. elements of delivery, such as direct training, network sessions, gap-tasks; or elements of content, such as emphasis on leadership of change; QA; classroom practice; marking and feedback) or does it work well as a package? • What are participants' suggestions for improving the programme? • What are participants' perceptions of the impacts of the Leadership Lite programme on leadership and teaching practices, teacher workload and satisfaction?	Head of science survey; teacher survey; case studies	<u>HoS and Teacher surveys (endpoint)</u>: C1 Summer 2021; C2 Summer 2023 <u>Case studies</u>: C1 Autumn 2019 and Summer 2021; C2 Autumn 2021 and Summer 2023
	• What do the Leadership Lite delivery team think about the quality and effectiveness of delivery? (e.g. quality of SLEs/consultants, quality of training and ongoing support and coordination from central Leadership Lite team, consistency of the quality of delivery across the clusters/SLEs/consultants, aspects of the programme that require improvement) • What factors are affecting the quality of delivery?	Developer interviews; observations; case studies (deliverer interviews with LL trainers/consultants/ SLEs)	<u>Developer interviews</u>: Autumn 2019, Autumn 2021, and Summer 2023 <u>Observations</u>: C1 Autumn 2019; C2 Autumn 2021 <u>Case studies</u>: C1 Autumn 2019 and Summer 2021; C2 Autumn 2021 and Summer 2023

Reach	• How many schools and staff are engaged in the programme? • How many schools/staff dropout during the programme and what are the reasons for this, if known?	MI data	<u>MI Data:</u> termly C1 and C2
Responsiveness	• How do school staff respond to Leadership Lite? (e.g. do they anticipate it will be useful? Are they convinced the strategies will support school improvement?) To what extent are the approaches advocated congruent with their existing beliefs about effective practice? Are they implementing the strategies? To what extent? • What additional support is required to engage with the Leadership Lite programme and its advocated school improvement practices? (e.g. do senior/middle leaders allocate additional time or support to implement the approaches?) • What are participants' views about the suitability, sustainability and potential for roll out of the intervention (e.g. are they likely to apply Leadership Lite practices to other areas of the school? Would they recommend that other schools participate in Leadership Lite? What is required to facilitate this?)	Head of science survey; teacher survey; case studies; observations	<u>HoS and Teacher surveys (endpoint):</u> C1 Summer 2021; C2 Summer 2023 <u>Case studies:</u> C1 Autumn 2019 and Summer 2021; C2 Autumn 2021 and Summer 2023 <u>Observations:</u> C1 Autumn 2019; C2 Autumn 2021
Programme differentiation	• To what extent do Leadership Lite participants think that the intervention is distinctive to practices prior to randomisation? • What practices and approaches does Leadership Lite replace in participating schools?	Case studies	<u>Case studies:</u> C1 Autumn 2019 and Summer 2021; C2 Autumn 2021 and Summer 2023
Monitoring of control group	• What are the business as usual practices in control schools (in relation to leadership, QA, classroom practice, marking and feedback) and how distinctive are these from Leadership Lite practices?	Head of science survey; teacher survey	<u>HoS and Teacher surveys:</u> C1 Spring 2019; C2 Spring 2021; C1 Summer 2021; C2 Summer 2023
Adaptation	• What challenges have schools faced in implementing the programme? • How have schools adapted Leadership Lite and why? (e.g. adapted their engagement with the support offered, such as involved different staff, or adapted their implementation of Leadership Lite practices) • What contextual factors have influenced implementation of the Leadership Lite practices? (e.g. senior leader support, time, amending policies, staff involved) • How have challenges been overcome?	Case studies	<u>Case studies:</u> C1 Autumn 2019 and Summer 2021; C2 Autumn 2021 and Summer 2023

	• What challenges have deliverers faced in providing the programme? (e.g. sustaining engagement, changes to staffing, drop out, supporting implementation in schools) • What adaptations will the developers make to the programme for Cohort 2 in light of learning from the first cohort?	Developer interviews; case studies (interviews with LL trainers/ consultants/ SLEs)	Developer interviews: Autumn 2019, Autumn 2021, and Summer 2023 Case studies: C1 Autumn 2019 and Summer 2021; C2 Autumn 2021 and Summer 2023

To gather this information, the implementation and process evaluation will employ the following methods:

Data collection with all trial schools

We will administer baseline and endpoint (online) surveys with Heads of science in intervention and control schools (140 schools). The online surveys will ask the same questions at baseline and endpoint to identify schools' practices at the starting point and changes to practice over time. The surveys will explore practice in relation to the leadership of the department's: workload; quality assurance of lessons; marking and assessment; and lesson planning and classroom practice. The survey will also ask about science staffing, including: whether there are any current science teacher vacancies; and, whether any science teachers have left in the last year and if so, why. At endpoint for intervention schools, the survey will include brief questions exploring perceptions of the quality and effectiveness of Leadership Lite training. For Heads of science in control schools, the endpoint survey will include an additional question about professional development activities undertaken during the intervention period. Each survey will take 10 minutes to complete.

We will administer baseline and endpoint (online) surveys with science teachers in intervention and control schools (140 schools). We will aim to receive responses from all eligible science teachers and deliver the survey via direct email (collected as part of the initial collection of teacher information from schools). The online surveys will ask the same questions at baseline and endpoint to identify schools' practice at the starting point and changes to practice over time. The surveys will explore attitudes and practice in relation to: school leadership; workload; quality assurance of lessons; marking and assessment; lesson planning and classroom practice; and satisfaction with, and engagement in, teaching. At endpoint for intervention schools, the survey will include brief questions exploring perceptions of the quality and effectiveness of Leadership Lite training. For teachers in control schools, the endpoint survey will include an additional question about professional development activities undertaken during the intervention period. The survey will collect a small amount of demographic data on characteristics that we expect to influence survey responses (for example, role, length of time teaching). Each survey will take 10 minutes to complete.

Schools will be required to provide science teacher details in order to be randomised. Baseline surveys will be administered prior to randomisation (i.e. Spring 2019 for cohort 1 and Spring 2021 for cohort 2) and endpoint surveys will be administered at the end of the intervention period for each cohort (i.e. Summer term 2021 for cohort 1 and summer term 2023 for cohort 2). The completion of the surveys is part of schools' participation agreement.

Data collection with all intervention schools

CET will collect termly Management Information (MI) from all intervention schools. This will comprise participation data for school staff in Leadership Lite training sessions, network events, gap tasks and online support. MI logs will also collect data on time and cost implications of participation with Leadership Lite. NFER will design the data collection logs in collaboration with CET for this purpose and agree submission dates. The MI returns will be completed termly by each school's senior leader key contact.

Data collection with a sub-sample of intervention schools

We will conduct case studies in six intervention schools (three in each cohort) to explore culture change in the implementation environment (such as the factors that have facilitated or impeded implementation of the Leadership Lite practices in schools, and the anticipated cultural, policy and practice changes identified in the intervention description and logic model above), as well as detailed insights on programme quality, fidelity and time/costs associated with engaging with the programme. The case studies will be selected to provide some variation in characteristics such as: location (to consider the effects of programme delivery within and beyond the immediate location of CET); science staff recruitment challenges; and school/science department development needs. We will seek CET's agreement to a proposed case-study sample selection.

We will conduct case-study interviews at the beginning and towards the end of each Cohort. The case studies will involve:

- telephone interviews with up to six local Leadership Lite trainers/consultants/SLEs working directly with the case-study schools, one trainer/consultant/SLE per case-study school

- observations of training (up to six training sessions, one observation per case-study school)

- telephone and face-to-face interviews with school staff, including: the headteacher/senior leader; a school governor; the head of science; and up to four science teachers at the beginning and end of their participation in the trial.

Interviews with Leadership Lite trainers/consultants will explore their experiences of the central training and coordination by Leadership Lite, perceptions of school engagement and response to the programme, how the programme is delivered in their cluster/s, and at follow-up perceptions of how the programme is supporting developments in schools.

Observations of training will explore how Leadership Lite is delivered, whether delivery appears to be consistent across different clusters, what is delivered and how closely it aligns with the recommended practices of Leadership Lite, as well as indications of quality, effectiveness and school response.

We will aim to interview staff in case-study schools who have varying characteristics, such as: subject taught, length of time teaching and role in the department. We will seek the assistance of the Head of science (or other key contact for the evaluation) in identifying and approaching interviewees. Initial case-study interviews with school staff will explore perceptions of: pre-planning and foundations for the programme; the implementation environment; reasons for applying for the trial; strategies previously adopted to encourage science teacher retention and/or recruitment; programme differentiation; and anticipated benefits and challenges. Interviews with science teachers will explore their current job satisfaction, what motivates/demotivates them; and whether they plan to stay (in the school, in teaching) or are considering leaving within the next year. Follow-up interviews with school staff will take place with the same individuals (wherever possible, or those occupying similar roles) and focus on: programme quality; implementation of support; fidelity and adaptation; key benefits/perceived outcomes, challenges and time/cost implications of engaging with Leadership Lite.

Data collection with developers

We will conduct telephone interviews with up to two representatives from CET at three time points during the course of the trial (i.e. at the beginning of the trial and delivery with Cohort 1, at the mid-point of the trial - end of Cohort 1 delivery and beginning of Cohort 2 delivery, and the end of the trial following Cohort 2 delivery). These provider interviews will focus on perceptions of the quality of delivery and school response, delivery challenges and key success factors, progress and adaptation, changes to the programme (particularly between Cohort 1 and 2).

Analysis of process evaluation data

We will triangulate all sources of data collected via the IPE to assess programme fidelity and non-compliance to the intervention. This data will be gathered via: MI data, developer interviews and observations; Heads of science surveys, teacher surveys and case studies. This will provide data on the extent and nature of engagement with the intervention. Anticipated compliance with the intervention is set out in the 'duration and dosage' section of the TIDieR framework above.

There are three aspects of dosage that the IPE will collect information on:

- the breadth of engagement within schools (e.g. the proportion of staff in a school engaged in Leadership Lite activities);
- the appropriateness of staff involved (e.g. the role of school staff involved, including governors, senior leaders, middle leaders and teachers);
- the extent of engagement (e.g. the number of Leadership Lite training sessions attended, the number of further support/implementation activities engaged with such as network sessions, gap tasks and online support).

A fidelity measure will be constructed for reporting and use in the on-treatment / compliance analysis outlined above. This will be defined in detail in the SAP, and is likely to be a school-level measure that draws on each school's extent of engagement with Leadership Lite. We will analyse the variation in fidelity across intervention schools and measure the proportion of intervention schools achieving an acceptable minimum level of compliance, pre-agreed with CET and defined in the SAP.

IPE data will be analysed using MAXQDA for qualitative data analysis and SPSS for quantitative analysis. Qualitative data from observations and interviews with developers, deliverers, and case-study schools will be summarised in write-up templates that explore key themes of the research questions (e.g. fidelity, quality of delivery, adaptation). Using MAXQDA, the data will be coded into these broader themes and sub-codes assigned to segments of text to enable detailed analysis of the content, range and prevalence of responses relating to each theme. Quantitative data from the Head of science and teacher surveys, and MI data will be analysed using SPSS. This analysis will be conducted with a full audit trail and quality-assured by a senior statistician. Analysis of survey data will primarily include comparing change over time in responses at baseline and endpoint to questions about workload, job satisfaction and a range of teaching and leadership practices for the intervention and control groups. Factor analysis will be conducted to create composite measures of survey questions that explore similar constructs. Sub-group analysis will be conducted to explore the relationship between variables of interest and changes to responses over time (e.g. teachers' length of time in teaching). The various data sources of

the IPE will be triangulated and reported in relation to each of the research questions/themes.

Cost evaluation

We will gather cost data by interviewing the programme developers to estimate the costs incurred to deliver the intervention, which schools would need to cover in the absence of EEF subsidy. Costs to schools, such as teacher time required to attend training and any resource costs incurred, will be measured through the MI data, endpoint survey with Heads of science, endpoint teacher survey and case study interviews. We will then combine these data sources into an estimate of the average cost per school and per pupil over three years, by separating one-off and on-going costs.

Ethics and registration

The trial will be designed, conducted and reported to CONSORT standards (http://www.consort-statement.org/consort.statement/) and registered on http://www.controlled-trials.com/. The evaluation will be conducted in accordance with NFER's Code of Practice, available at: http://nfernet/media/3029/code_of_practice_final_january_2019.pdf. NFER, CET and EEF will work together to ensure each organisations' policies can be applied in practice.

Ethical agreement for participation within the trials will be provided by the headteacher or the senior leader of the school via signing an MoU. Teachers in the science department will be provided with full details about the intervention, and will be given the opportunity to withdraw their data processing if they have objections to this.

Data protection

All data gathered during the trial will be held in accordance with the data protection framework created by the Data Protection Act 2018 and the General Data Protection Regulation 2016/679 , and will be treated in the strictest confidence by the NFER, CET and EEF. No individual or school will be identified in any report. Teacher data collected from schools by NFER will not be made available to anyone outside of the parties listed. Our legal basis for gathering and using this data is our legitimate interest to administer the randomised controlled trial and analyse its data.

NFER and CET are the joint data controller for this evaluation.

The legal basis for processing personal data is covered by:

GDPR Article 6 (1) (f) which states that 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'. We have carried out a legitimate interest assessment, which demonstrates that our legitimate interest for processing personal data for this trial is to administer the RCT and analyse its data.

NFER,CET and EEF are signing a Data Sharing Agreement that sets out the roles and responsibilities for this trial. This includes a description of the nature of the data being collected and how it will be shared, stored, protected and reported by each party. In addition, CET will provide a memorandum of understanding to schools, explaining the nature of the data being requested of schools, teachers and pupils, how it will be collected, and how it will be passed to and shared with NFER.

For the purposes of the trial, CET will collect name, role and contact details of a key contact person at schools when they are recruited. CET will share this data with NFER via NFER's secure data portal. NFER will collect head of science and teacher details directly from schools using NFER's secure data portal for schools. This will include names, job title/role, TRN, date of birth, school and personal contact email address, length of time in teaching, subject taught and degree of specialism. NFER will also administer questionnaires to these individuals, undertake observation of Leadership Lite training, conduct telephone interviews with CET, intervention providers, school senior leaders, governors, heads of science and science teachers. NFER will share teacher administrative data with DfE in order to access the data held in the SWC. The data sharing will take place via DfE's secure data exchange portal.

Within three months of the end of project, NFER will send school and teacher data to EEF's data archive partner. This will include school names, administrative teacher data (such as teacher name, date of birth, TRN and role) matched to SWC, data from MI logs and surveys. At this point, EEF's data archive partner will keep a copy of the data and EEF will become the Data Controller. Anonymised data will also be stored with the DfE, the Office for National Statistics (ONS) and potentially other research teams. Further matching to SWC and other administrative data may take place during subsequent research.

NFER and CET will retain personal data for one year after report publication in case there are any queries about the report. One year after report publication, all personal data will be securely deleted.

Personnel

David Bailey (Director of Research & Development, CET) will be the project lead, taking responsibility for oversight of the programme.

Jack Worth (Lead Economist, NFER) will be the Project Director, taking responsibility for project delivery and quality assurance. Palak Roy (Senior Trials Manager, NFER) will lead and manage the trial on a day-to-day basis, act as a daily contact with CET and EEF, and oversee the impact evaluation. Caroline Sharp (Research Director, NFER) will direct the process evaluation, supported by Jennie Harland (Research Manager, NFER). NFER's Research and Product Operations department will oversee school communications and data collection from schools, including administering staff surveys.

Risks

Risk	Assessment	Countermeasures and contingencies
Insufficient schools recruited to the study	Likelihood: **moderate** Impact: **high**	NFER could help with recruitment for an additional fee if this becomes problematic.
School or teacher attrition	Likelihood**: low** Impact: **moderate**	Clear information/initial meeting with schools explaining the principles of the trial and expectations plus MOU. Trial design using administrative data limits potential for attrition.
Intervention is not implemented well	Likelihood: **low** Impact: **moderate**	Clear information/initial meeting with schools explaining the principles of the trial and expectations. Both 'intention to treat' and 'on-treatment' analysis will be used. Good communication with delivery team to provide strong implementation. Process evaluation will monitor implementation.
Control group adopts similar treatments (contamination issues)	Likelihood: **moderate** Impact: **moderate**	CET committed to preventing Cohort 1 control schools from applying to participate in Cohort 2. Process evaluation will monitor extent and nature of contamination in control group.
Differences between cohorts	Likelihood: **moderate** Impact: **moderate**	The team will capture differences between cohorts through MI data, provider interviews, case studies and surveys.
Unable to access SWC and NPD data	Likelihood: **very low** Impact: **very high**	On-going NFER and EEF engagement with DfE over third-party data access arrangements.

Timeline

Dates	Activity	Staff responsible/ leading
July - Aug 2018	Set up meetings 1 and 2.	EEF/ NFER/ CET
Sep 2018 – Jan 2019	IDEA workshop Write protocol	NFER/ CET
Nov 2018 - Apr 2019	Cohort 1 recruitment and schools sign MoU Schools provide list of science teachers	CET
Feb – May 2019	Baseline teacher data collection in all Cohort 1 schools Baseline survey to heads of science and science teachers in all Cohort 1 schools	NFER
May 2019	Randomisation Inform schools of randomisation outcome	NFER CET
Sept – Dec 2019	Cohort 1 delivery begins Recruit case study schools and arrange visit Observe training and conduct baseline case-study interviews Baseline interviews with developers Carmel Education Trust provides termly MI data	CET NFER
Jan - Dec 2020	Match teachers to SWC 2018 and 2019 Carmel Education Trust provides termly MI data	NFER CET
Nov 2020 - April 2021	Cohort 2 recruitment and schools sign MoU Cohort 2 schools provide list of science teachers Carmel Education Trust provides termly MI data (Cohort 1)	CET
May – Aug 2021	Endpoint survey to heads of science in Cohort 1 (intervention and control) Endpoint survey to science teachers in Cohort 1 (intervention and control) Carmel Education Trust provides termly MI data Baseline survey to heads of science and science teachers in all Cohort 2 schools Endpoint telephone interviews with Cohort 1 case study schools and providers Midpoint interviews with developers Cohort 1 delivery ends	NFER CET NFER CET
May 2021	Cohort 2: Randomisation Inform schools of randomisation outcome	NFER CET
Sept - Dec 2021	Cohort 2 delivery begins Recruit case study schools and arrange visit Observe training and conduct baseline case-study interviews Carmel Education Trust provides termly MI data	CET NFER
Jan – Dec 2022	Match teachers to SWC 2020 and 2021. Carmel Education Trust provides termly MI data	NFER CET
Jan - Dec 2023	Endpoint survey to heads of science in Cohort 2 (intervention and control) Endpoint survey to science teachers in Cohort 2 (intervention and control) Carmel College provides termly MI data Endpoint telephone interviews with Cohort 2 case study schools and providers Endpoint interviews with developers Cohort 2 delivery ends	NFER CET NFER

Jan - Aug 2024	NPD request NPD data analysis for Cohorts 1 and 2 Match teachers to SWC 2021 and 2023 SWC and data analysis for Cohorts 1 and 2 IPE data analysis	NFER
Sept – Dec 2024	Draft 1 report to EEF EEF review and team to revise report	NFER / EEF
Jan - Mar 2025	Draft 2 report for peer review Final report revisions, submit final report and sign off	NFER / EEF

References

Adey, P. and Shayer, M. (1994). Really raising standards: cognitive intervention and academic achievement. London, Routledge.

Bailey, D.M. (forthcoming) Reducing teacher workload in science using the Leadership Lite approaches.

Black, P. and Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. Phi Delta Kappa 80 (2) 139-144, 146-148.

Coe, R., Aloisi, C., Higgins, S. and Major, L.E. (2014). *What Makes Great Teaching?* Available [online]: https://www.suttontrust.com/wp-content/uploads/2014/10/What-Makes-Great-Teaching-REPORT.pdf (8 November 2018).

Department for Education (2017) *Teachers Analysis Compendium: 2017.* Available [online]: https://www.gov.uk/government/statistics/teachers-analysis-compendium-2017 January 2019).

Driver, R., Squires, A., Rushworth, P. and Wood-Robinson, V. (1994) *Making sense of Secondary Science: Research Into Children's Ideas.* London: Routledge

Goleman, D (2000). Leadership that gets results. *Harvard Business Review*, Vol. 78 (2).

Hattie, J. and Timperley, H. (2007). 'The power of feedback', *Review of Educational Research,* 77, 81.

Hattie, J. C. (2009) *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London & New York: Routledge, Taylor & Francis Group

Hill, A., Mellon, L., Laker, B. and Goddard, J. (2016a). *The One Type of Leader Who Can Turn Around a Failing School.* Available [online]: https://hbr.org/2016/10/the-one-type-of-leader-who-can-turn-around-a-failing-school (8 November 2018).

Hill, A., Mellon, L., Goddard, J. and Laker, B. (2016b). *How to Turn Around a Failing School.* Available [online]: https://hbr.org/2016/08/how-to-turn-around-a-failing-school (8 November 2018).

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) *Implementation and process evaluation (IPE) for interventions in education settings.* Available [online]: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Guidance_Final.pdf (16 January 2019).

Joyce, B. and Showers, B. (2002). *Student Achievement Through Staff Development*, 3rd Ed. Alexandria, VA, ASCD.

Public Accounts Committee. (2018). *Retaining and developing the teaching workforce.* Available [online]: https://www.parliament.uk/business/committees/committees-a-z/commons-select/public-accounts-committee/inquiries/parliament-2017/sustaining-developing-teaching-workforce-17-19/ (30 January 2019).

Quigley, A., Muijs, D. and Stringer, E. (2018) *Metacognition and Self-regulated learning: Guidance report.* Available [online]:

https://educationendowmentfoundation.org.uk/tools/guidance-reports/metacognition-and-self-regulated-learning/ (30 January 2019).

Sadler, D. R. (1989) Formative Assessment and the Design of Instructional Systems, *Instructional Science*, vol. 18, pp. 119-44

Shute, V. (2008) 'Focus on formative feedback', *Review of Educational Research* 78 (1) p153.

Timperley, H., Wilson, A., Barrar, H. and Fung, I. (2007). *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration.* Available [online]: http://www.oecd.org/education/school/48727127.pdf (8 November 2018).

Worth, J. and De Lazzari, G. (2017) *Teacher Retention and Turnover Research - Research Update 1: Teacher Retention by Subject*. Available [online]: https://www.nfer.ac.uk/teacher-retention-and-turnover-research-research-update-1-teacher-retention-by-subject/