Trial Evaluation Protocol: SMART SPACES

Evaluator: UCL Institute of Education

Principal investigator: Jeremy Hodgen

Template last updated: March 2018

PROJECT TITLE	SMART Spaces: Spaced Learning Revision Programme
DEVELOPER (INSTITUTION)	Queen's University Belfast / Hallam Teaching School Alliance
EVALUATOR (INSTITUTION)	UCL Institute of Education
PRINCIPAL INVESTIGATOR(S)	Jeremy Hodgen
PROTOCOL AUTHOR(S)	Jeremy Hodgen, Jake Anders, Nicola Bretscher, Mark Hardman
TRIAL DESIGN	Randomised Controlled Trial, two-arm design, efficacy trial: intervention vs. business as usual control
PUPIL AGE RANGE AND KEY STAGE	15 – 16 year olds in Year 11 studying double award science, KS4
NUMBER OF SCHOOLS	125
NUMBER OF PUPILS	12500 (approximately 100 pupils per school)
PRIMARY OUTCOME	GCSE Science (Double Award) Chemistry Sub-Score
SECONDARY OUTCOME	AQA GCSE Double Award Science; Knowledge, application and analysis assessment objectives sub-scales for the Chemistry element of AQA GCSE Double Award Science

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.2		
1.1		
1.0 [*original*]	xx	*[leave blank for the original version]*

Table of contents

Abstract

This evaluation protocol describes the evaluation of SMART Spaces: Spaced Learning Revision Programme (SMART Spaces Revision), an efficacy trial funded by the Education Endowment Foundation (EEF), designed to investigate the effect of the intervention on the chemistry element of the GCSE double award science. The evaluation will consist of a two-arm randomised controlled trial testing the SMART Space Revision intervention against a 'business as usual' control and will include a mixed methods implementation and process evaluation (IPE). The trial will take place over the 2018/19 academic year, with randomisation taking place in two stages in October and December 2018 and the delivery of the intervention in schools during April and May 2019, following teacher training and coaching sessions over the period December 2018 to March 2019. Final publication of the results will be in Summer 2020. This protocol outlines the rationale for the project, describes the intervention using the TIDieR framework and outlines the methods of data collection and analysis for the impact evaluation, the IPE and the cost evaluation.

Introduction

This evaluation protocol describes the design and methods for the evaluation of SMART Spaces: Spaced Learning Revision Programme (SMART Spaces Revision), an efficacy trial funded by the Education Endowment Foundation (EEF), to investigate the effect of the intervention on the chemistry element of the GCSE double award science. The intervention, SMART Spaces Revision, is developed by a team from Queen's University Belfast (QUB) and Hallam Teaching School Alliance (HTSA) [the developer]. The evaluation will be carried out by a team from the UCL Institute of Education (UCL) [the evaluator].

The evaluation will consist of a two-arm randomised controlled trial testing the SMART Space Revision intervention against a 'business as usual' control and will include an embedded mixed methods implementation and process evaluation (IPE). The trial will take place over the 2018/19 academic year, with randomisation taking place in two stages in October and December 2018 and the delivery of the intervention in schools during April and May 2019, following teacher training and coaching sessions over the period December 2018 to March 2019.

This trial follows a promising evaluation of a pilot study, also funded by the EEF (O'Hare, Stark, McGuinness, Biggart & Thurston. 2017). This is one of two concurrent evaluations; the other, a pilot of a teaching version of SMART Spaces, is described in a separate evaluation protocol (Hodgen, Anders, Bretscher & Hardman, In prep).

Intervention

The intervention is a further development of the original pilot, designed to be delivered at scale across the 50 schools in the intervention arm. The short description that follows is based on the *Template for Intervention Description and Replication* (TIDieR) checklist , which should be read in conjunction with the logic model (Figures 1a and 1b). A SMART Spaces manual will provide further guidance for teachers and schools.

1. *Brief name*: SMART Spaces Chemistry Revision Version

2. *Why (rationale/theory)*: An educational programme for AQA GCSE double (or combined) award science students used for examination revision in chemistry using spaced learning. Evidence from the neuroscience and cognitive psychology indicates that including spaced intervals between learning sessions can improve factual recall, and an earlier pilot suggested that a combination of short (10 minute) and longer (approximately 24 hour or period of sleep) spaces provides a promising model of spacing (see O'Hare et al., 2017). Therefore it is the spacing that is of fundamental importance in this intervention. It is anticipated that factual recall will impact on the application and analysis as well as knowledge elements of the chemistry score in GCSE double award science.

3. *Who (recipients)*: Year 11 pupils in schools across England.

4. *What (materials)*: PowerPoint chemistry revision slides covering the entire GCSE double science chemistry curriculum content to be used in intervention spaced lessons. SMART Spaces manual, and SMART Spaces activity pack to be used by teachers. Materials for spacing activities during intervention spaced lessons (e.g. juggling balls).

5. *What (procedures)*: The intervention consists of two elements: SMART Training (delivered by SMART trainers, who are recruited and themselves trained) and SMART Spaces Implementation.

SMART Training: All teachers who will deliver SMART Spaces are trained in a half-day training session, which includes a demonstration of part of a SMART Spaces lesson and a chance to try out delivery in a trial run and get initial feedback. All schools will receive a manual and video at the training session, which exemplifies the approach and explains the importance of fidelity.

As part of the SMART Training coaching visits will take place between January and April 2019. During coaching visits, each teacher is observed delivering at least a 15-20 minute segment of a SMART Spaces lesson, including teacher delivery and transition to/from a space. These will likely be with examination classes (affecting dosage – some pupils will receive more than the prescribed number of iterations of the material). Trainers will have a standard pro forma to complete to support provision of feedback to teachers who have been trained. Feedback will be provided either face to face or by phone and will cover timing, content delivery, evidence of teacher preparation, classroom organisation and transitions.

We anticipate that some teachers may further practice spacing activities prior to intervention (affecting dosage).

SMART Spaces implementation: Pupils are prepared for the intervention in a prior lesson by the teacher explaining that the following revision lessons will use the SMART approach.

Chemistry topics for AQA Paper 1 are taught in the three short ~12-minute sessions: A, B and C, with 10-minute spaces between each topic; A-B and B-C. Additional spacing is assumed to occur before and after the lesson (-A and C-) due to changes in activity. Where SMART Spaces lessons take place in the second part of a double lesson, there should be a short sensorimotor activity to separate any teaching of content from the initial spaced materials being delivered. The spacing involves a sensorimotor activity from a menu of suitable activities including juggling. The developers will provide schools with juggling balls and advise them that juggling was used in the pilot study and found to be feasible in the classroom. They will also provide a list of alternative activities and advise schools they may choose from this list if they do not feel juggling is practical within their lessons. These are:

- Balloon games (e.g. two lines of students racing to pass balloon from front to back)
- Modelling clay
- Origami
- “Simon Says”

This full lesson (A-B-C) is delivered three times on three separate days over a minimum of three days and a maximum of a week (thus providing additional spaces of around 24 hours (at least) between content repetitions, during which pupils sleep). For compliance, all three parts of the paper 1 materials, and two spaces must be delivered within a lesson, and there must be at least one sleep before repetition of the lesson. After at least one further sleep, but ideally the following week, the process is repeated for content associated with AQA Paper 2: content D, E and F.

6. <u>*Who (implementers)*</u>: The SMART Spaces intervention lessons are delivered by GCSE science teachers who have had SMART Training. The same teacher should provide the whole SMART Spaces programme, ideally in two consecutive weeks within the three-week intervention window.

SMART Training and Coaching is provided by trainers experienced in the delivery of SMART Spaces. Heads of Science will also be present at SMART training to ensure that departmental implementation is coordinated and supported, and that the SMART Spaces intervention is integrated within the school's overall approach to science revision.

7. *How (mode of delivery)*: A whole-class programme that is conducted during six normal science lessons. SMART Training is delivered to groups of teachers. SMART Coaching is delivered to teachers by trainers in one-to-one sessions following observations of a practice SMART session.

8. *Where (setting)*: SMART Training conducted in out-of-school (or twilight in-school) sessions, SMART Coaching takes place in school, and SMART Spaces intervention in standard GCSE classroom.

9. *When and how much (dosage)*: The programme of six intervention spaced lessons covers AQA GCSE chemistry curriculum content in a high intensity way and is delivered in two blocks of three days during the three week period prior to the first GCSE double award examination. The SMART Spaces slides are set out in six 12-minute chunks of GCSE chemistry content (approximately one sixth of each course) to be taught in one-hour lessons, the same lesson delivered on three days (A-B-C x3), then a second, different, lesson on a further three days (D-E-F x3). There is an expectation that a teacher's delivery of the 12 minute chunks becomes more efficient over the three consecutive days, as less elaboration takes place in repetitions. Pupils experience six one-hour periods in total.

For this trial, Tuesday 23rd April – Fri 10th May inclusive is the window in which the intervention groups must undergo 3 iterations of Paper 1 content (A-B-C), followed by 3 iterations of Paper 2 content (D-E-F). This is based on the draft exam timetable for summer 2019 which indicates that Chemistry paper 1 is scheduled for Thurs 16th May, and Chemistry paper 2 for Wednesday 12th June 2019.

10.*Tailoring*: SMART Spaces Chemistry Revision is a manualised intervention and treatment fidelity should be maximised. Teachers can choose from a menu of spacing activities. It is expected that teachers will become more efficient over the three iterations of A-B-C, and likewise three iterations of D-E-F. This is due to teacher learning, less verbal embellishment by the teacher and better recall by students. This may allow for some adaptation of the time spent on particular topics and the provision of more feedback to pupils.

Some teachers may share slides with pupils (although this will not be actively encouraged). Pupils engaged in the intervention may therefore use the slides and/or adopt spacing within their own revision practices.

11.*How well (planned)*: Effective implementation requires training all teachers who teach Y11 GCSE chemistry in all trial schools before they deliver the intervention spaced lessons. This training will consist of modelling, practice, and feedback on programme delivery. It is anticipated that teacher enthusiasm will influence the fidelity and quality of delivery of the intervention spaced lessons as well as whether the SMART revision lessons are delivered. Effective implementation also requires support from a Head of Science to ensure that lesson are scheduled and supported within the revision period.

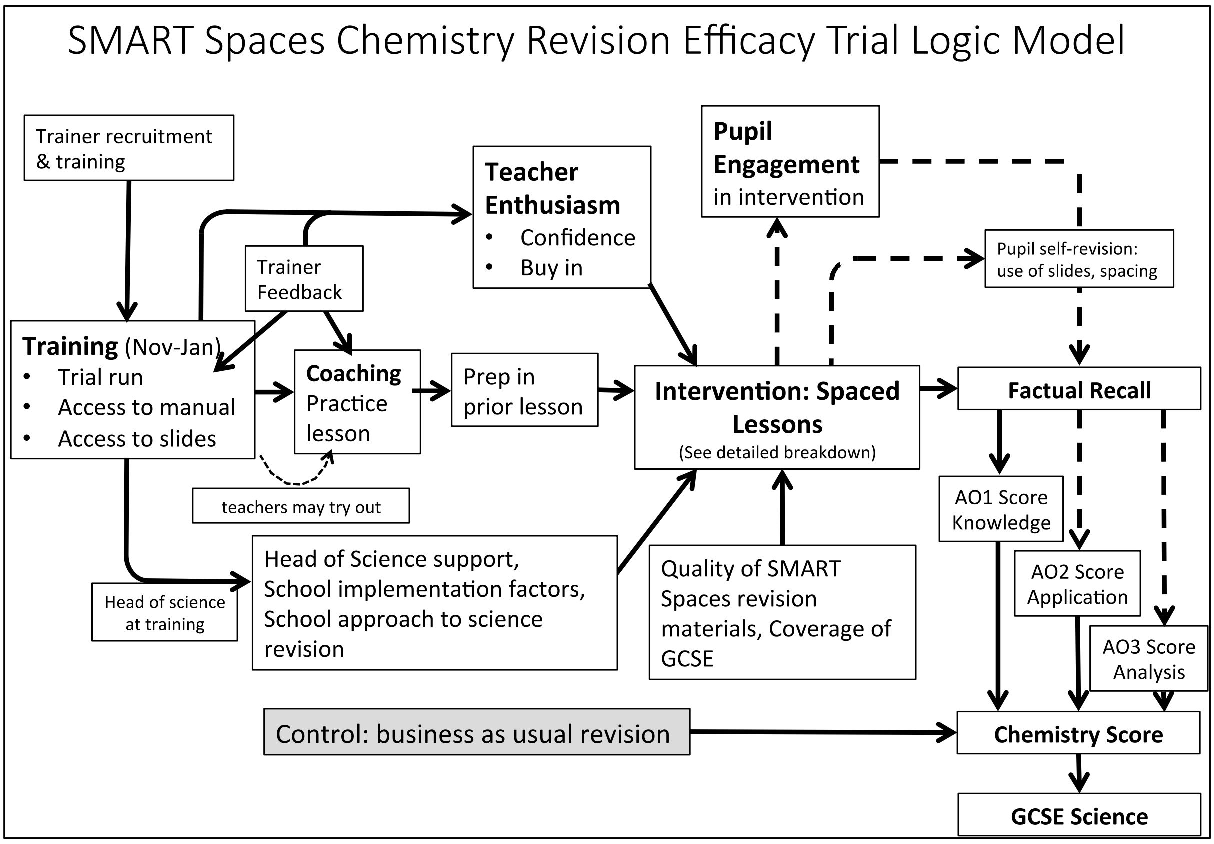

Figure 1a: SMART Spaces Chemistry Revision Logic Model (overall)

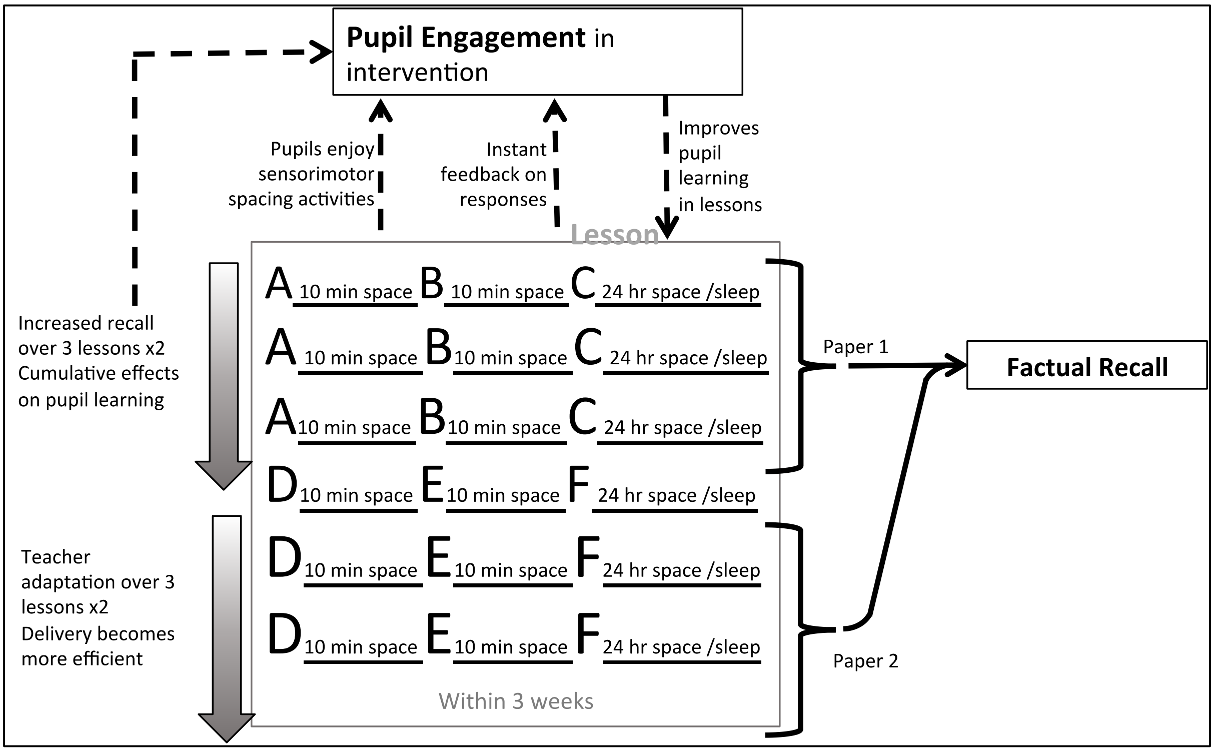

Figure 1b: SMART Spaces Chemistry Revision Logic Model (intervention spaced lessons element)

Study rationale and background

Spaced learning is a promising development for science education (with important implications for other subjects, such as mathematics). The key theoretical principle of a spacing effect is that repetitions of a learning activity are more effective for learning when the repetitions are separated by a space in time rather than clustered together in a massed format. It is likely that a long space between repetitions is even more effective for improving learning when it includes a period of sleep (Bell et al., 2014), such as the 24 hour space between SMART Spaces lessons.

Short spaces (10 minutes) have also been found to be effective in the neuroscience literature (Fields, 2009) although less is known about the role a particular activity could play within a short space for improving learning. For obvious reasons sleep is not possible for short spaces yet a distractor task is necessary to minimise the chance of verbal or mental rehearsal of the material which would stop the space between learning periods actually being a break from explicit learning activity. A physical task which requires a lot of conscious attention on physical activity is therefore the natural choice for a distractor task to both provide a contrast to the auditory and visual input of the SMART Spaces materials and to reduce the chance of students consciously rehearsing the material and shortening the space between periods of learning. All spacing activities on the menu meet these criteria.

Neuroscience research has found that physical activity increases synaptic long-term potentiation in humans (Smallwood et al., 2015) and improves language learning (Winter et al., 2007). It is possible then that the most physically active task, juggling, may offer the most advantage for learning. Furthermore, rapid horizontal eye movements (saccadic eye movements) and tactile stimulation alternating between the two hands in the retention period between learning and recall have also been found to improve recall (Nieuwenhuis et al., 2013). It is thought that rapidly alternating activation of both hemispheres is beneficial for memory formation. Both of these potential facilitators of memory formation are present during juggling (saccadic eye movements when tracking balls and alternating tactile stimulation when catching). Rapid eye movements are performed more by novice jugglers than by expert jugglers as skilled juggling often includes gaze fixation (Dessing et al., 2012), so it is possible that the alternating hemisphere activation benefit to memory may reduce as juggling skill improves. For these reasons, juggling may offer some additional benefits to memory formation, but this is theoretical, and the key criteria for spacing activities is that they contrast with the learning periods, discourage discussion and mental rehearsal of the material and that they involve a physical activity.

The theoretical benefits of these spacing activities should be considered within the context that they have not been subjected to direct experimental study of their effect on recall. The primary facilitator of change in memory formation is the robust and frequently demonstrated spacing effect - the advantage of spaced presentation over massed presentation.

In the pilot evaluation, O’Hare et al. (2017) provide an informative review of the evidence highlighting that, whilst the neuroscience and cognitive psychology literature indicate a robust spacing effect, the mechanisms underlying the spacing effect are poorly understood and there are several competing theories of how spacing affects learning (see, e.g., Smolen, Zhang and Byrne, 2016). Hence there is a need for applied research to address the large gap between the research evidence and classroom practice.

A significant question relates to the optimal length and organisation of spacing. The original pilot study addressed this question by conducting a feasibility study (FS) and an optimisation study (OS) examining the effects of short spaces (10 minute), long spaces (24 hour) and a combination of the two compared to an active control (learning the same material without spacing) and a business as usual control. The different length spaces tested in the OS both utilised interleaving (ABC-ABC-ABC, rather than AAA-BBB-CCC), and O'Hare et al (2017, p. 35) note that a nuanced understanding of the effects of this (Taylor and Rohrer, 2010) have yet to be explored in relation to SMART Spaces. The most promising approach in the OS was the combination of 10 minute and 24 hour learning which showed a statistically significant ES of g=0.19 compared to the business as usual control with a smaller non-significant ES of g=0.11 compared to the active control using a bespoke test based on GCSE examination questions across science. Moreover, although the OS suggested that the potential of spaced learning appears to be greater for short answer questions, which tend to focus on recall, this 10 minute / 24 hour combination variant appeared to show some promise on longer answer questions, which place more emphasis on synthesis and application of knowledge. In addition, the OS provided evidence of the feasibility and readiness of the approach. However, we lack firm evidence of the intervention working in schools at scale. As such, we believe there is the appropriate equipoise to conduct a randomised controlled trial to estimate the impact and provide new data on what works to improve academic attainment for pupils in English schools.

Impact Evaluation

Research questions

The evaluation will address the following primary research question:

RQ1. What is the size of the effect of the SMART Spaces intervention on pupils' attainment in the chemistry element of GCSE 'double award' science when compared to a business-as-usual control, and is the effect practically distinguishable from a null effect?

In addition, the evaluation will address the following secondary research questions:

RQ2. What is the size of the effect of the SMART Spaces intervention on pupils' attainment in GCSE 'double award' science when compared to a business-as-usual control, and is the effect practically distinguishable from a null effect?

RQ3. What is the size of the effect of the SMART Spaces intervention on pupils' attainment in the assessment objectives constituting the chemistry element of GCSE 'double award' science (knowledge, application and working scientifically) when compared to a business-as-usual control, do the size of the effects differ for the difference objectives, and are the effects practically distinguishable from a null effect?

RQ4. What is the size of the effect of the SMART Spaces intervention on the attainment of pupils eligible for free school meals compared to other pupils, and is the effect practically distinguishable from either a null effect or the effect on pupils' attainment in general?

RQ5. Are the effects on attainment practically distinguishable for girls and boys?

Additional questions relating to the Implementation and Process Evaluation (IPE) are discussed below.

Design

The trial is designed as school-level randomised controlled trial in schools offering double award science through the AQA board involving approximately 12,500 pupils in 125 schools. It will be a two-arm efficacy trial: SMART Spaces Chemistry Revision compared to a business as usual control. The developer has limited capacity to deliver the training and coaching to schools. In order to ensure the trial has sufficient power given this limited capacity, allocation to the arms will be unequal. We expect to allocate 50 schools to the intervention and 75 to the business as usual control. See Figure 2 for an overview of the trial design and timeline.

Table 1: The trial design

Trial type and number of arms		Two-arm, cluster randomised
Unit of randomisation		School
Stratification variables (if applicable)		Randomisation block, School-level prior attainment
Primary outcome	variable	Chemistry attainment
	measure (instrument, scale)	Chemistry sub-scale of AQA GCSE Double Award Science (item-level, continuous)
Secondary outcome(s)	variable(s)	[1] Science attainment [2] Knowledge, application and analysis elements of chemistry attainment
	measure(s) (instrument, scale)	[1] AQA GCSE Double Award Science (UMS, continuous) [2] Knowledge, application and analysis assessment objectives (AO) sub-scales for the Chemistry element of AQA GCSE Double Award Science (item-level, continuous)

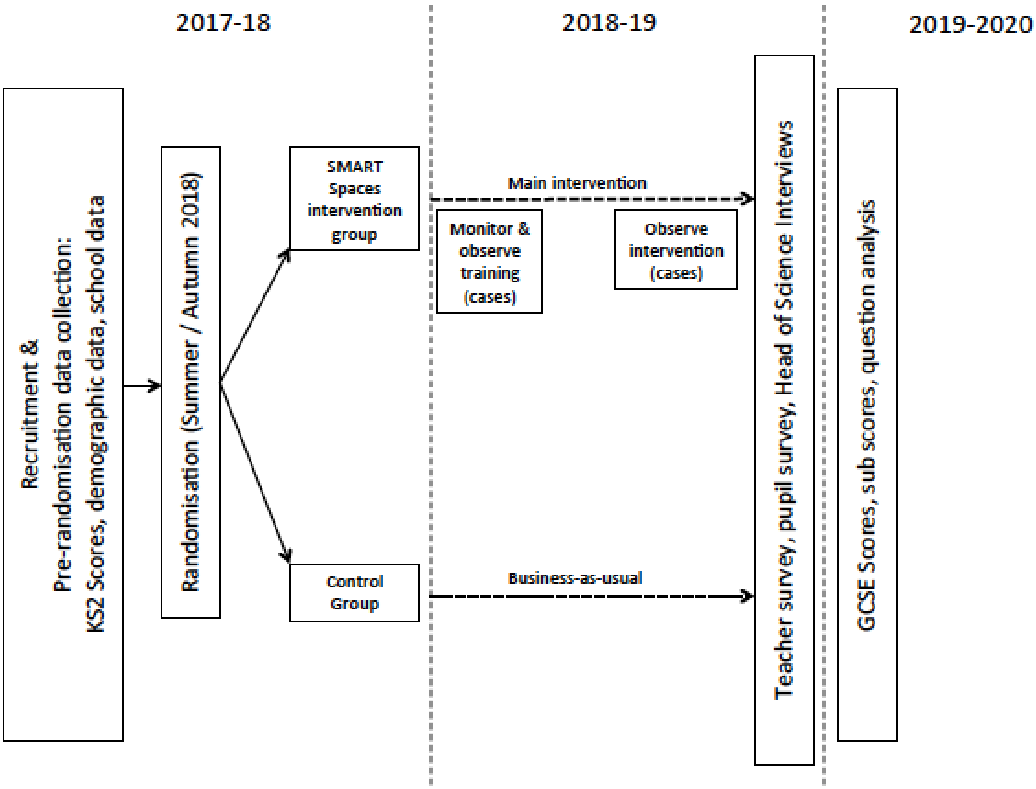

Figure 2: Overview of the trial design

Randomisation

Randomisation is planned take place in two batches, one no later than October 2018, and one in December 2018, with a target of 90 schools in the first randomisation batch. This is to aid recruitment and delivery. If recruitment proceeds more quickly than expected and it is possible to randomise all participating schools together in October this will be done.

We will use a school-level randomisation approach. Our preferred approach would be to do simple randomisation using a four blocked design stratified by:

1. randomisation batch (2 groups), and
2. prior attainment of pupils (2 groups).

Hence. Randomisation will be in two batches, each blocked by attainment. Since the objective is to maximise internal validity by achieving sufficient balance between the two arms of the trial, the attainment blocks will be defined by a median split based on sample characteristic (ie school average attainment). However, randomisation in batches can introduce imbalance. We will simulate randomisation in advance in order to assess the relative merits of different approaches, in particular whether further blocking is necessary to avoid imbalance between two arms, or whether a minimisation procedure is preferable. This simulation will be carried out shortly before the first randomisation date based on the actual sample of schools. A further simulation will be carried out prior to the second andomisation date.

Pupil-level randomisation was considered, but, although this would increase power, this was judged likely to be unacceptable to many schools in the crucial GCSE year, and, in addition, the potential for within-school contamination of the revision approach would be high.

Randomisation and the statistical analysis will carried out by Nicola Bretscher under guidance by Jeremy Hodgen and Jake Anders. In order to ensure that analysis was blind to group allocation, the dataset will be blinded by another member of the evaluation team.

Participants

Schools offering the double (or combined) award GCSE science through the AQA board will be eligible for the trial. Recruitment will be led by the developer and will focus initially on the following areas (expanding the recruitment regions if necessary): North of England, and the East and West Midlands. If recruitment proves problematic, an option to relax eligibility criteria to allow schools offering double award GCSE science through other exam boards will be considered. If so, we will examine the feasibility of creating a robust chemistry sub-scale across awards from the different examination boards. Recruitment will aim for the overall proportion of FSM pupils at school-level to be as high as or higher than the average across England.

The trial will involve only those Y11 pupils who are entered for the double (or combined) award GCSE science. One of the findings of the OS was that teachers considered the intervention to have greater benefits for low attaining students and that higher attaining students were less engaged and perceived there to be less benefit (O'Hare, 2017, p.32). Hence, the intervention may have a smaller impact on highest attaining students, and restricting the target group to students enrolled for double award GCSE science was judged to be a way of increasing power (because the triple award is taken by higher attaining pupils). Schools will not participate in another EEF GCSE science randomised trial that would interfere with implementation of the intervention with Year 11 pupils during 2018/19 academic year.

Recruitment is led by QUB from the developer. Initially, schools in the North of England and the East and West Midlands were approached via either a mailshot, direct email or via school networks (including multi-academy trusts, teaching schools, local authorities and science learning partnerships). In order to increase recruitment, a second mailshot expanded the regional recruitment focus. Schools expressing interest were provided with a detailed briefing pack, and followed up by email and, where necessary, telephone. To be included in randomisation, schools will have to provide a Memorandum of Understanding signed by the school's head teacher (and constituting the school's agreement to take part in the project) outlining the responsibilities of the school, the developer and the evaluator. Specifically, this requires schools to:

- commit to the outcome of the randomisation process,
- confirm GCSE Science examination board and specification , and,
- return student information for all Year 11 pupils enrolled in double award science (student first and surnames, date of birth, gender, Unique Pupil Number (UPN), eligibility for FSM, KS2 mathematics and English scores, students science class and teacher, confirm entry to double award science)

Schools in the control group will receive £1000 following the completion of all evaluation requirements with staff/school and with the required pupils in 2018 and 2019. After the evaluation has finished, the school may purchase the SMART Spaces programme from QUB/HTSA for use from January 2020.

Sample size calculations

Table 2: MDES Calculations: Overall and everFSM

		OVERALL	FSM
MDES		0.198	0.231
Pre-test/ post-test correlations	level 1 (pupil)	0.50	0.50
	level 2 (school)	0.25	0.25
Intracluster correlations (ICCs)	level 2 (school)	0.15	0.15
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Average cluster size		100	25
Number of schools	Intervention	50	50
	Control	75	75
	Total	125	125
Number of pupils	Intervention	5000	1250
	Control	7500	1875
	Total	12500	3125

This trial faced two issues in determining the sample size and estimating power. First, as noted above, the developer has limited capacity and is able to deliver the training and coaching to a maximum of 50 schools. Second, the correlations between Key Stage 2 national test results, the proposed pre-test, and GCSE double (or combined) award science are around 0.55 at pupil-level (Benton & Sutch, 2014). In order to ensure the trial would have sufficient power, a number of options for unequal allocation to the intervention and control groups were considered. These power calculations were carried out using the R package, PowerUpR. They share the standard assumptions of 0.8 power for a two-tailed significance test at the 0.05 level; intra-cluster correlation is assumed to be 0.15; there are assumed to be four blocks and 100 pupils in each school; and a single pre-test regressor (KS2 English and mathematics combined) is assumed with a correlation with the outcome measure (GCSE chemistry sub-score) of 0.5 at pupil-level and 0.25 at school-level. The bases for these assumptions are shown in Table 3.

This suggests that an MDES of 0.198 can be achieved using a sample of 125 schools unequally allocated to intervention and control in a ratio of 50:75. (See Table 2.) To allow for some attrition a recruitment target of 135 was agreed (with an absolute maximum of 54 schools allocated to the intervention group). For the FSM subgroup analysis a conservative estimate of 25 pupils per school suggests that an MDES of 0.231 can be achieved.

Table 3: Justification of assumptions for power calculations

Assumption	Justification
ICC = 0.015	Based on EEF (2015) ICC document.
100 pupils per school	This is equivalent to around four Double Science GCSE classes per school. Since around 75% of pupils take the double award[1] and the average size of a secondary school is around 180 pupils per year group, this is a relatively conservative assumption.
Pupil-level pre- to post-test correlation = 0.5	No data are available for the correlation between the chemistry element of the GCSE double award science and Key Stage 2 (KS2) scores. The correlation was estimated on the basis of the correlation between GCSE science and combined KS2 scores of 0.556 (Benton & Sutch, 2014). This was judged a better predictor than the correlation between triple award chemistry and KS2. However, since this is lower (0.427), a conservative estimate was judged appropriate.
School-level pre- to post-test correlation = 0.25	This was estimated to be half the pupil-level correlation on the basis of advice from the EEF evaluation advisory panel.

Outcome measures

Primary Outcome:

(i) Chemistry sub-scale of AQA GCSE Double Award Science, a continuous numerical variable based on item-by-item mark data

Secondary Outcomes:

(ii) AQA GCSE Double Award Science (based on Uniform Marking Scale (UMS) scores rather than numerical grades)

(iii) The knowledge, application and analysis assessment objectives (AO) sub-scales for the Chemistry element of AQA GCSE Double Award Science, a continuous numerical variable based on item-by-item mark data

The Chemistry sub-scale (based on item-by-item mark data) has been chosen as the primary outcome to maximise the power of the trial by avoiding dilution of the impact on chemistry attainment with attainment in the other sciences.[2]

With regard to the secondary outcome, this has been chosen to measure the impact on GCSE science as a whole. However, GCSE grades are not designed to form a linear scale. In order to simplify the modelling and increase discrimination, we will use the Uniform Marking Scale (UMS) score, which can be modelled using linear regression techniques. The

[1] https://ffteducationdatalab.org.uk/2017/03/weird-science/

[2] The dilution effect was estimated to be between 13% and 33% (i.e., the effect on GCSE double award science as a whole would be between 13% and 33% of the effect on the chemistry element alone.

UMS score is a tool that all exam boards use to standardise marks awarded on papers across the different exam boards and paper tiers.

The knowledge, application and analysis assessment objectives (AO) sub-scales have been chosen to investigate whether the intervention can have an impact on application and analysis skills, as well factual recall. The theoretical basis of spaced learning suggests that there would be greater benefits for factual recall. Further support for this is provided by the pilot evaluation in which greater effects were found for short answer items in comparison to long answer items.

The Chemistry sub-scale, UMS and item-by-item mark data will be collected directly from schools in August and September 2019. We will conduct a dummy run of data collection in 2018 using pilot (non-trial) schools identified by the developers (HTSA) in order to develop guidance for schools on downloading and submitting this data, and to pilot the coding of the Chemistry knowledge, application and analysis (AO) sub-scales and the overall chemistry scores. We will additionally collect numerical GCSE science grades, which will be used to impute data if missing data is high. In the unlikely event that it does not prove possible to collect UMS scores, we will use numerical grades as a primary outcome. Since the trial uses national test data for both pre- and the primary and secondary outcomes post-tests, there is no need to ensure the testing is further blinded. Bespoke pre- and post-tests were considered. However, these were judged likely at best to produce pre-to-post tests correlations no greater than between the GCSE chemistry sub-scale and KS2, and, hence, to have no advantage over national tests in terms of power, whilst being less cost effective and more burdensome to schools.

Pre-test measures: A simple aggregation of English and mathematics KS2 scores, noting the importance of literacy for science attainment (Nunes, 2017), and the increased emphasis on mathematics within the science GCSE, (OFQUAL, 2015).

Analysis plan

All quantitative outcomes would be modelled on the basis of intention to treat (ITT) using a linear multi-level model. We propose to fit a 2-level multi-level model of students clustered in schools. All models will be estimated using Bayesian inference (Gelman et al., 2014) with the software STAN through R using weakly informative and diffuse priors.

We will specify an acceptable level of attrition on the basis of simulation of the dataset and amend the protocol when the Statistical Analysis Plan is agreed following randomisation. Results will be modelled on the basis of intention to treat (ITT). Our analysis will be based on a 2-level model incorporating the treatment condition, and the pre-test and other stratification variables used for randomisation as covariates.

Effect sizes will be calculated using the Hedges g ES for cluster randomised trials as per the current EEF (2018) statistical analysis guidance for evaluations. The primary and secondary outcomes will be reported using 95% Bayesian credible intervals. We will also report classical confidence intervals to enable comparability with other EEF trials. In a Bayesian framework, there is no direct equivalent to null hypothesis testing. Following Kruschke and Liddell (2017), we will use a ROPE (Region of Practical Equivalence) analysis to examine whether the null hypothesis should be accepted as credible or practically distinguishable. This procedure examines the proportion of the Highest Density Interval (HDI) that falls within a pre-determined effect size. Although, this effect size is often set at ±0.1 around 0 in Bayesian analyses, it is important to consider what constitutes the most appropriate level to set taking account of two factors: comparability with other EEF trials reported using a standard frequentist approach, and the size of effect that is judged likely to be 'practically significant' for pupil learning.

Additional sensitivity analyses will be conducted (e.g., replicating the analysis with different software). These analyses will be detailed in the statistical analysis plan (SAP), which will be produced after randomisation.

Subgroup analysis

A sub-group analyses will be carried out for everFSM pupils (defined as any pupil who has ever been classified as in receipt of free school meals). In addition, because the under-participation of girls in science is judged to be an important issue for both policy and research (Royal Society, 2014; TISME, 2013), a sub-group analysis will be carried out for sex. For the sub-group analyses, we will add an interaction effect to the main primary effects model. If the interaction is credible in Bayesian terms, we will run a separate model using only the relevant sub-group. This analysis will require the primary outcome data to be matched with the National Pupil Database (NPD), to provide 'everFSM' data for pupils. NPD data will be matched to original pupil data collected before randomisation.

Secondary outcomes

We will model the secondary outcomes as separate models (rather than through a multivariate multilevel model) given that they consist of questions from the GCSE chemistry sub-scale. Like with the overall score, we will model on the basis of intention to treat, with KS2 combined score included as a pre-test.

Implementation and process evaluation

A robust and in-depth implementation and process evaluation (IPE) is vital to ensure we understand how the SMART Spaces intervention is implemented, and the extent to which the logic model (see Figures 1a and 1b) adequately describes the factors and mechanisms underlying the intervention as well as the key conditions for success and any barriers to implementation. Our IPE will take a mixed methods approach. We outline the RQs, data collection and analysis below.

Research Questions

In the process evaluation, we will address the following research questions [*Italicised comments* in square brackets cross-reference each question to Humphrey et al's (2016) "*Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook*"]:

RQ6. Was SMART Spaces implemented with fidelity in the trial, and to what extent can SMART Spaces be implemented with fidelity in a scaled-up version of the intervention? *[Fidelity, implementation]*

RQ7. Are there any barriers to implementation? *[Fidelity, adaptation, implementation factors]*

RQ8. What role do heads of science play in facilitating implementation? *[Fidelity, adaptation, implementation]*

RQ9. What are the most and least effective aspects of training teachers to deliver SMART Spaces with fidelity? *[Quality, (teacher) responsiveness, implementation]*

RQ10. Do teachers and heads of science, perceive SMART Spaces to be a useful and engaging approach to revision? *[Quality, (teacher and school) responsiveness, reach]*

RQ11. To what extent does teacher engagement affect the quality of delivery and pupil responsiveness? *[Quality, (teacher and pupil) responsiveness, implementation]*

RQ12. (a) Do teachers trial the spaced lessons before the intervention and do they practice or prepare in any other way? (b) Do they adopt spaced learning in other chemistry revision lessons? *[Dosage]*

RQ13. (a) To what extent do teachers adapt the materials and approach? (b) In what ways do teachers and schools adapt their approach to science revision as a result of SMART Spaces? *[Adaptation]*

RQ14. (a) Are all pupils responsive to SMART Spaces and does it have reach: do all pupils perceive it to be an engaging and beneficial approach to revision? (b) What contributes to pupil engagement (or disengagement)? *[Reach]*

RQ15. Do some pupils adopt spacing practice within their own revision practices? *[Reach, differentiation, dosage]*

RQ16. To what extent is SMART spaces distinguishable from 'business as usual' revision practice in schools? *[Differentiation, monitoring of control group, implementation]*

RQ17. To what extent does the logic model (see Figures 1a and 1b) adequately describe the mechanism by which the SMART Spaces intervention effected change (if any)? [*Differentiation, intervention characteristics*]

Implementation and process evaluation data collection

Data collection will involve questionnaires and surveys, case studies and interviews as set out below. A central aim of the IPE is to evaluate the extent to which the logic model describes the key factors and mechanisms of the intervention and it implementation. Hence, in Appendix 1, we set out how these data are linked to the logic model.

Questionnaires and Surveys: Questionnaires and surveys will be short and use either optical character recognition (OCR) or online technology. We propose to collect the following data for both the intervention and active control groups:

- Survey to be completed by the head of science/head of chemistry in each intervention school, administered online (with a paper-based option) to gather data on their perceptions of the intervention and its implementation, how implementation was supported, and the quality and variability of delivery. The survey will be piloted and tested on a small sample of heads of science before use. *[Fidelity, (schools) responsiveness, adaptation, reach, differentiation, implementation factors, RQ6, RQ7, RQ8, RQ9, RQ11, RQ12, RQ13, RQ17].* (n_{HoSInt}~50)[3]

[3] Numbers in parentheses indicate sample numbers. Response rates are difficult to predict and cannot be guaranteed, but previous experience suggests a response rate of 60% might be expected.

- Survey to be completed by the head of science/head of chemistry in each business as usual control school, administered online (with a paper-based option). (n_{HoSCon}~75). We would also administer a survey to 2 teachers of chemistry within each control school, in addition to the head of science/chemistry (n_{TCon}~150). These would investigate schools' revision practices, and whether schools use spacing (or interleaving) approaches (and thus allow us to better understand the counterfactual) *[differentiation, monitoring control group, RQ16, RQ17].*

- Survey to be completed by all teachers in intervention (n_{TInt}~400), administered online (with a paper-based option) to gather data on their perceptions of the intervention and its implementation, how implementation was supported, the quality and variability of delivery and teacher perceptions of student engagement. The survey will be piloted and validated on a sample of teachers before use. *[Fidelity, (teacher) responsiveness, adaptation, reach, quality, differentiation, implementation factors, RQ6, RQ7, RQ8, RQ9, RQ10, RQ11, RQ12, RQ13, RQ17].*

- Survey of all treatment students (n_{SInt}~7500), adapted from the implementation survey used in the pilot to gather data on student engagement, revision practices more generally in school science, and their approaches to revision at home. We will validate this survey more fully using statistical techniques (e.g. Rasch modelling) and further piloting. Additional items will be added to gauge school revision practices, which will be informed by previous work on student ratings of instruction (e.g. Nitz et al., 2014) *[Fidelity, dosage, (students) responsiveness, reach, adaptation, RQ6, RQ14, RQ15].*

- Survey of a sample of students in the control schools (n_{SCon}~150), to establish current revision practices in schools and their approaches to revision at home. We will validate this survey more fully using statistical techniques (e.g. Rasch modelling) and further piloting. *[Monitoring control group, RQ16].*

<u>*Case Studies*: </u>Given the relatively complicated nature of the training and intervention, and the multifaceted nature of the logic model, we will use case studies in order to elucidate the processes and evaluation. *[Fidelity, (teacher) responsiveness, adaptation, reach, quality, implementation factors, RQ6, RQ7, RQ8, RQ9, RQ11, RQ12, RQ13, RQ14, RQ17]*

We will conduct 5 case studies of school implementation, each primarily following 3 teachers, as well as the Head of Science/Chemistry. Schools will be selected purposefully in order to cover a range of levels of school engagement with the SMART Spaces revision programme. Data collected will include:

- Observation of SMART training sessions (n=5)

- Observation of teacher feedback in practice (coaching) sessions, and brief interview with trainers (n=15)

- Observation of 2 of the 3 intervention spaced lessons (2x(A-B-C) or 2x(D-E-F)) . We will observe the whole lesson where the timetable allows, and, in other cases, will observe parts of lessons. At least one entire lesson, and parts of at least two lessons, will be observed for each teacher. (n=30)

- Brief interviews with each case study teacher (n=15)

- Interviews with Head of Science (n=5)

Observations will follow a pre-determined protocol, and interviews will be semi-structured.

Interviews (Control): In order to better understand the business as usual control, particularly revision practices, in addition to the surveys described above, we will conduct short telephone interviews with Heads of Science/Chemistry in control schools (n=5).

Additionally interviews with the developers (QUB, HTSA) in order to better understand the intervention and thus inform the IPE data collection.

Implementation and process evaluation data analysis

Questionnaire and survey data: Surveys will be analysed descriptively and, where appropriate comparisons can be made, using inferential statistics. If the measures of student engagement are judged to be sufficiently robust, we will explore the effect of student engagement quantitatively through interaction analysis using the models from the impact evaluation.

Case Study Data: The case study data and interviews will be analysed thematically (e.g. Braun & Clarke, 2006) informed by the survey results.

Analysis of Materials: In order to further assess quality we will also analyse the revision material relative to the AQA chemistry specification, with a particular focus on coverage of curriculum. [*Quality,* programme differentiation, *RQ12, RQ16.*]

Assessment of usual practice: We will assess 'usual' practice at baseline via the Head of Science and teacher surveys for intervention and control schools, and in case studies for intervention school. Given the focused and well-defined nature of this intervention (revision in run up to GCSE), we anticipate that teachers will have reliable recall of revision practices in previous years and, hence, that the end point surveys will be sufficient to capture baseline practice.

Table 4 provides a summary of the how the data and RQs relate to the IPE dimensions and implementation factors.

Table 4: Overview of how data addresses IPE dimensions, factors and research questions

IPE Dimension / Factor	RQs	Data
Fidelity	RQs 6, 7 & 8 &9?	Intervention group surveys: Head of science; teacher; pupil. Case study: training & coaching observation, lesson observation; interviews with head of science, teachers & SMART trainers. Developer records of training / coaching.
Dosage	RQs 12 & 15	Intervention group surveys: Head of science, teacher & pupil. Case study: interviews with head of science.
Quality	RQs 9, 10 & 11	Intervention group surveys: Head of science, teacher & pupil. Case study: lesson observation; interviews with head of science & teachers. Analysis of materials.
Reach	RQs 14 & 15	Intervention group surveys: Head of science, teacher & pupil. Case study: lesson observation; interviews with head of science & teachers.
Responsiveness	RQs 9, 10, 11 & 14	Intervention group surveys: Head of science, teacher & pupil. Case study: lesson observation; interviews with head of science & teachers.
Programme differentiation (and the assessment of ‘usual practice’ at baseline and endpoint)	RQs 15, 16 & 17	Intervention group surveys: Head of science & teacher. Control group surveys: Head of science &teacher. Case study: lesson observation; interviews with head of science, teachers & SMART trainers.
Monitoring of control group	RQ 16	Control group surveys: Head of science & teacher, pupil.
Adaptation	RQs 7, 8 & 13	Intervention group surveys: Head of science & teacher. Case study: lesson observation; interviews with head of science & teachers.
Preplanning & foundations	RQs 6, 7 & 8	Intervention group surveys: Head of science & teacher; pupil. Case study: interviews with head of science & SMART trainers. Developer records of training / coaching.
Implementation support system	RQs 6, 7, 8 & 9	Intervention group surveys: Head of science & teacher. Case study: training & coaching observation; interviews with head of science, teachers & SMART trainers.
Implementation environment	RQs 6, 7 & 8	Intervention group surveys: Head of science & teacher. Case studyinterviews with head of science.
Implementer factors	RQs 6, 7, 8 & 11	Intervention group surveys: Head of science & teacher. Case study: lesson observation; interviews with head of science & teachers.
Intervention characteristics	RQ 16 & 17	Intervention group surveys: Head of science & teacher. Case study: lesson observation; interviews with head of science & teachers.

Non-compliance analysis

Compliance will be analysed at school-level using an Instrumental Variables (IV) approach with group allocation as the instrumental variable for the compliance indicator. The definition of minimum compliance will be based on attendance at training and coaching sessions and the delivery of SMART Spaces lessons. These will be agreed with the developer by September 2018. Attendance registers from training and coaching sessions will be collected from the developer in order to assess attendance, and teachers and students will be asked to about the number of SMART Spaces lesson delivered through the surveys above (allowing us to triangulate these data).

We will also investigate the effects of “non-compliance” in the control group. The SMART Spaces intervention is not publically available, so schools in the control group will not have access to the intervention materials. However, there may be some schools, or teachers, in the control group who use a spaced learning approach for revision, and we will attempt to capture these “always compliers" using survey data, and if sufficiently robust data are available, we will investigate control group non-compliance quantitatively.

Cost evaluation

We will follow the June 2016 EEF Guidance on Cost Evaluation in estimating the costs of the delivery of the intervention. Costs will be reported as an average cost over three years per double award student who is in receipt of SMART Spaces lessons. With regard to the direct, marginal costs, we will estimate the costs of providing SMART training, coaching and resources by the developer, and any costs incurred by schools to cover travel to training (if required) and any additional resources. We will collect cost data from the developer via a short interview and either a pro-forma or developer records. In addition, we will collect data on costs incurred by schools through the process evaluation (through case studies and Head of Science surveys). In addition to staff time to attend training, we will estimate the staff time required to plan, implement and support SMART Spaces using evidence collected during the process evaluation, using both survey data from teachers (which can tend to overestimate the time required for new interventions of this sort) and data from the case studies . As per the EEF guidance, we will report ‘time’ required separately to other costs. We anticipate that the bulk of the costs will be incurred during the first year of implementation, but we will discuss with the developer whether any additional costs would be incurred on additional training or coaching (e.g., for new staff) or to replace resources. If required we will estimate these costs through the IPE.

Ethics and registration

The trial has had approval from the relevant ethics committees of both UCL and QUB: UCL IOE Research Ethics Committee Reference: REC1052. QUB Research Ethics approved 11/04/2018 by SSESW, QUB Research Ethics Committee.

We intend to process personal data for public interest purposes. (See data protection below.) Neverthless, we will provide an opportunity for parents/ carers and pupils to withdraw their own, or their child’s data, from any data processing as part of the research to ensure that they have no objection to their data being processed in this way. This will demonstrate that the processing does not impinge on anyone’s rights and meet our responsibilities under the BERA Ethical Guidelines for Educational Research (particularly regarding informed consent, openness and disclosure).

Parents, and participating pupils, will be informed of the research through information sheets distributed by schools, along with withdrawal forms to support the process described above. The information sheets and withdrawal forms for this purpose explain the intervention and the research being conducted in simple language, provide opportunities for parents to ask additional questions, and provide clear steps to follow if they wish their child to be withdrawn from any data processing as part of the research. The sheet and form also make it clear that data can be withdrawn at this point or at any point up to 31st August 2019, in line with requirements to ensure participation is free from coercion.

The Implementation and Process Evaluation (and validation of the associated instruments and protocols) involves more active participation of teachers and pupils, including lesson observation and interviews. To this end, we propose to collect unambiguous consent from participating teachers, the parents and carers of participating pupils and the pupils themselves. Information sheets and consent forms for this purpose are included with this application.

If information that raises safeguarding concerns is raised by a teacher or pupil during their discussions with us we will liaise with the relevant school's safeguarding officer regarding the appropriate course of action. Our information sheets make clear that disclosures of this type cannot remain confidential and will be reported. The researchers carrying out these interviews understand the need to manage disclosure carefully and sensitively. If in doubt, they will request advice from a senior colleague.

Outcomes of the project will be publicly reported through an EEF evaluation report and subsequent academic publications. No outcomes will include reporting that could allow for the identification of schools or pupils that participated in the research. The impact estimates will be reported as aggregated statistics while the implementation and process evaluation reporting will ensure that any references to individual schools, teachers and pupils are anonymised or removed, where residual risk of identification remains. Impact evaluation data will be securely shared with the EEF's Data Archive (managed by FFT) as part of their strategy for long term follow-up.

The trial will be registered with the ISRCTN (www.controlled-trials.com) following publication of this evaluation protocol.

Data protection

Data will be processed in line with data protection legislation (including the General Data Protection Regulation, GDPR), and in line with the interests of the participants. The project is registered with the UCL Data Protection Officer (registration number: Z6364106/2018/03/25 social research). Each organisation has carried out an assessment of their legal basis for processing data. Data will be processed by UCL and QUB on the basis of the public task purpose (as per condition 6(1)e of the GDPR), and by HTSA on the basis of the legitimate interest purpose (as per condition 6(1)f of the GDPR). UCL has reviewed current ICO guidance available here: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/, and has determined that this research forms part of its performance of a task in the public interest, as one of its core purposes provided for in its Charter and Statutes. (See Appendix 2 for a statement of the lawful basis and public tasks assessment for data processing).

We do not believe that any of the data we process falls within the definition of special category data under the GDPR. This would require an additional justification under Article 9(2) of the GDPR.

Pupils and their parents or carers, and teachers, will be informed of the proposed data processing and given an opportunity to object to this, and withdraw their, or their child's, data. The information which will be provided to parents/carers, pupils and teachers explains in clear and plain non-technical language the purpose to which we will put the data, that they can object to this data and this will be respected, contact details of the organisation, and categories of data that we will be processing and that the data processing will be compliant with the GDPR and data protection legislation. Further details on the lawful basis for data processing are available on request.

The evaluation team at UCL have carried out a data protection impact assessment and will put in place a data management plan. As part of this data management plan, data will be checked and cleaned to ensure the GDPR principle (d) of accuracy is met.

Data security

All personal data collected or obtained as part of this project will be treated as "Highly Restricted" under UCL Data Protection classification guidance. Personal data (pupil names, UPNs, dates of birth, FSM eligibility, sex, national test results, class and teacher, as well as teacher names and survey data) will be stored, processed and analysed on the UCL Data Safe Haven (DSH), the technical infrastructure that UCL has built specifically to host sensitive research data.

Qualitative data will be pseudonymised. Once pseudonymised it will be stored in a secure folder on the UCL network within a project folder only accessible to project team members (using appropriate access control methods), and the pseudonymisation key stored on the DSH. Fieldnotes and audio recording will be stored in a locked filing cabinet within a locked office at UCL to which only the SMART Spaces research team will have access.

Some data transfer will be required between collaborators on this project at UCL and QUB. This will be conducted by making a secure remote connection (e.g. VPN) to between the university networks and transferring data across this. In addition, the data will be encrypted before sharing using a password shared between research team members by separate communication.

Schools will be required to submit personal data to UCL. This will be conducted via the Data Safe Haven's direct data transfer portal. Schools will be provided with clear guidance on securely submitting and protecting this data.

Online surveys for teachers will be administered through UCL's REDCap survey system whereby data is uploaded directly to the DSH in an encrypted form.

A risk assessment has been conducted for the storage, processing and transfer of all personal data for the SMART Spaces project. All team members undertake regular annual data security training.

The DSH environment is certified to ISO27001:2013 with BSI – certificate number: IS 612909. The most recent external audit was in May 2017. The hosting is on a thin client system (DSH) with dual factor authentication. This is a multi-user system with permission-based access control. The DSH is subject to penetration testing on an on-going basis. The DSH has its own firewall separating it from the UCL corporate network and the UCL network has a corporate firewall with a default deny policy for inbound connections. The DSH remote access mechanism is protected by a SSL certificate issued by Terena as well as DualShield dual factor authentication, which couples an Active Directory password with token-based authentication. Connections are AES256 encrypted. Data is transferred into the DSH system via a secure gateway technology which uses SSL/TLS with data retained via policy and systems that prevent data leakage.

Data will be kept for at least the duration of the project, until successful submission of the data to the EEF's data archive has been agreed by the funder. We may keep anonymised data beyond this period for the purpose of supporting submissions and revisions to submissions to academic journals. They will be kept for no longer than 10 years in line with UCL's guidance on retention of records for research.

UCL and QUB will sign a data sharing agreement outlining data security and protection issues.

Personnel

QUB and HTSA Development and Delivery Team:

Dr Liam O'Hare (QUB): SMART Spaces Co-designer and overall project direction
Alastair Gittner (HTSA): SMART Spaces Co-designer and training lead
Dr Patrick Stark (QUB): SMART Spaces Project Manager
Dr John Coats (HTSA): Director of Hallam Teaching School Alliance and HTSA lead
Dr Maria Cockerill (QUB): Recruitment Manager and school contact lead
Professor Alan Thurston (QUB): Expert Advisor
Professor Carol McGuinness (QUB): Expert Advisor
Ewan MacRae (QUB): PhD Student, Teacher CPD
Research Fellow (QUB - TBA): Fieldwork, analysis & contact with schools

UCL Institute of Education Evaluation Team:

Professor Jeremy Hodgen: PI, overall direction and impact evaluation lead.
Dr Jake Anders: Advice on the impact evaluation and statistical techniques.
Dr Nicola Bretscher: will undertake the statistical analysis under guidance from Hodgen and Anders, and will contribute to all other aspects of the evaluation.
Dr Mark Hardman will lead the IPE and will contribute to all other aspects of the evaluation.
Research Officer (TBA): IPE fieldwork and analysis & contact with schools.
Administrator (TBA): day-to-day support to the project, including supporting data collection.

Risks

Table 5 outlines an assessment of the potential risks associated with this evaluation and the action proposed to address them.

Table 5: Risk assessment for the evaluation

Risk	Likelihood	Impact	Action
Failure to recruit	Low / Moderate	High	• Establish timeline for recruitment involving a variety of methods • Regular developer and evaluator team contact
Failure to gain GCSE chemistry sub-score and item-by-item data from schools	Moderate	High	• Use NPD as back-up • Dummy run to develop guidance in Summer 2018
Attrition of schools	Moderate	Moderate / High	• Over-recruit schools for efficacy trial (target 135 schools) • Financial incentives for control schools • Regular contact with intervention and control schools • Allocate staff time to school liaison at key data collection points • Regular developer and evaluator team contact
Loss of staff	Low / Moderate	Low	• UCL IOE has a large staff team and would reallocate staff
Fidelity / compliance to control conditions	Moderate	Low / Moderate	• Monitor through regular contact with schools.
Poor response rate to pupil and teacher surveys	Low / Moderate	Moderate	• Monitor through regular contact with schools. • Regular developer and evaluator team contact • Financial incentives for control schools

Timeline

Table 3 outlines the main activities of associated with the delivery and evaluation of the SMART Spaces intervention. A more detailed timeline is provided in Figure 3.

Table 3: Timeline of the main activities

Dates	Activity	Responsible/ leading
Mar-Nov 2018	Recruitment	QUB
May-Jun & Sept-Oct 2018	Initial data collection, pre-randomisation	UCL
Oct & Dec 2018	Randomisation (two stages)	UCL
Dec 2018 - Mar 2019	SMART Spaces training	HTSA
Jan-April 2019	SMART Spaces coaching	HTSA
23rd April-10th May 2019	SMART Spaces intervention delivery	HTSA
16th May & 12th June 2019	GCSE science examinations	-
Aug-Sept 2019	Data collection (UMS scores)	UCL
Feb 2020	Submit draft final report	UCL

SMART Spaces (Revision Model) Evaluation: Timeline

Phase	Task	Responsibility	Mar'18 - Dec'18	Jan'19 - Dec'19	Jan'20 - Jul'20
Intervention Delivery	Recruitment	QUB / HTSA	■■■■■■■■□□	□□□□□□□□□□□□	□□□□□□□
	Resource Preparation	QUB / HTSA	■■■■■■■■□□	□□□□□□□□□□□□	□□□□□□□
	Training / CPD	HTSA	□□□□□□□□■■	■□□□□□□□□□□□	□□□□□□□
	Coaching session (in school / class)	HTSA	□□□□□□□□□□	■■■□□□□□□□□□	□□□□□□□
	Intervention delivery [*]	HTSA	□□□□□□□□□□	□□□■■□□□□□□□	□□□□□□□
Procedures	Ethics	UCL / QUB	■■□□□□□□□□	□□□□□□□□□□□□	□□□□□□□
	Data Sharing Agreements	UCL / QUB	■■■■■■■■■□	□□□□□□□□□□□□	□□□□□□□
	Data Security Training	UCL / All	■■□□□■■■□□	□□□□□□□□□□□□	□□□□□□□
	Data Transfer Information for Schools	UCL	□□□■■□□□□□	□□□□□□□□□□□□	□□□□□□□
	Publication Protocol	UCL / QUB	□□□□■□□□□□	□□□□□□□□□□□□	□□□□□□□
Evaluation	Evaluation Protocol	UCL	■■■■■■■■■□	□□□□□□□□□□□□	□□□□□□□
	Trial Registration	UCL	□□□□□□□□□■	■□□□□□□□□□□□	□□□□□□□
	Initial data collection (UPNs etc)	QUB / HTSA / UCL	□□□□□□■■■□	□□□□□□□□□□□□	□□□□□□□
	Randomisation (Two stages) [***]	UCL	□□□□□□□■□■	□□□□□□□□□□□□	□□□□□□□
	UMS Dummy Run	UCL / HTSA	□□□□□□□■■■	■□□□□□□□□□□□	□□□□□□□
	Pilot coding of science GCSE	UCL / QUB / HTSA	□□□□□□■■□□	□□□□□□□□□□□□	□□□□□□□
	Check & clean initial data	UCL	□□□□□□■■■■	□□□□□□□□□□□□	□□□□□□□
	Statistical Analysis Plan (SAP)	UCL	□□□□□□□□□□	■■□□□□□□□□□□	□□□□□□□
	Develop Obs Protocol (Training)	UCL	□□□□□□□■□□	□□□□□□□□□□□□	□□□□□□□
	Develop Obs Protocol (Coaching)	UCL	□□□□□□□□□■	□□□□□□□□□□□□	□□□□□□□
	Develop Obs Protocol (Revision sess)	UCL	□□□□□□□□□□	■■■□□□□□□□□□	□□□□□□□
	Develop Interview Schedules	UCL	□□□□□□□□□□	□□□■■□□□□□□□	□□□□□□□
	Develop / Validate Surveys	UCL	□□□□□□□□□■	■■■□□□□□□□□□	□□□□□□□
	Identify Case Study Schools	UCL	□□□□□□□□□□	■□□□□□□□□□□□	□□□□□□□
	Observation (Training)	UCL	□□□□□□□□■■	■□□□□□□□□□□□	□□□□□□□
	Observation (Coaching)	UCL	□□□□□□□□□□	■■■□□□□□□□□□	□□□□□□□
	Observation (Revision sessions) [*]	UCL	□□□□□□□□□□	□□□■■□□□□□□□	□□□□□□□
	Developer interviews	UCL / QUB / HTSA	□□□□□□□□□□	■■■■□□□□□□□□	□□□□□□□
	Additional Case Study Data (Interviews)	UCL	□□□□□□□□□□	■■■■■□□□□□□□	□□□□□□□
	Pupil survey (Intervention)	UCL	□□□□□□□□□□	□□□□■□□□□□□□	□□□□□□□
	Teacher survey (Intervention)	UCL	□□□□□□□□□□	□□□□□■■□□□□□	□□□□□□□
	Head of Science survey (Intervention)	UCL	□□□□□□□□□□	□□□□□■■□□□□□	□□□□□□□
	Head of Science Interviews (Intervention)	UCL	□□□□□□□□□□	□□□□□■■□□□□□	□□□□□□□
	Pupil survey (Control)	UCL	□□□□□□□□□□	□□□□■□□□□□□□	□□□□□□□
	Teacher survey (Control)	UCL	□□□□□□□□□□	□□□□□■■□□□□□	□□□□□□□
	Head of Science survey (Control)	UCL	□□□□□□□□□□	□□□□□■■□□□□□	□□□□□□□
	Head of Science Interviews (Control)	UCL	□□□□□□□□□□	□□□□□■■□□□□□	□□□□□□□
	GCSE [**]	Schools	□□□□□□□□□□	□□□□■■□□□□□□	□□□□□□□
	NPD Application	UCL	□□□□□□□□□□	□□□■□□□□□□□□	□□□□□□□
	Data collection (UMS scores)	QUB / HTSA / UCL	□□□□□□□□□□	□□□□□□□■■□□□	□□□□□□□
	Analysis (IPE)	UCL	□□□□□□□□□□	□□■■□□■■■■■■	□□□□□□□
	Analysis (Impact)	UCL	□□□□□□□□□□	□□□□□□□■■■■■	■■□□□□□
	Write Report	UCL	□□□□□□□□□□	□□□□□□□□□□■■	□□□□□□□
	Submit Draft Final Report	UCL	□□□□□□□□□□	□□□□□□□□□□□■	□□□□□□□
	Reviewer Feedback & Re-draft	UCL	□□□□□□□□□□	□□□□□□□□□□□□	■■□■□■□
	Report to Developer	UCL	□□□□□□□□□□	□□□□□□□□□□□□	□□□□■□□
	Report published	UCL	□□□□□□□□□□	□□□□□□□□□□□□	□□□□□□■

* Intervention delivery window: 23/4/19 - 10/5/19

** GCSE Science Provisional Dates (AQA): 14/5/19 (Combined Science Synergy Paper 1); 16/5/19 (Combined Science Trilogy Chemistry Paper); 22/5/19 (Combined Science Synergy Paper 2)

Figure 3: A detailed timeline of the evaluation activities

References

Bell, M. C., Kawadri, N., Simone, P. M., & Wiseheart, M. (2014). Long-term memory, sleep, and the spacing effect. Memory (Hove, England), 22(February 2015), 276–83. http://doi.org/10.1080/09658211.2013.778294

Benton, T., & Sutch, T. (2014). *Analysis of use of Key Stage 2 data in GCSE predictions. Report commissoned by Ofqual Ref:14/5471*. Coventry: Office of Qualifications and Examinations Regulation.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Dessing, J. C., Rey, F. P., & Beek, P. J. (2012). Gaze fixation improves the stability of expert juggling. Experimental Brain Research, 216(4), 635–644. http://doi.org/10.1007/s00221-011-2967-6

EEF. (2015). *Intra-cluster correlation coefficients* Retrieved from https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/ICC_2015.pdf

EEF. (2018). *Statistical analysis guidance*. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

Fields, R. D. (2009). Making memories stick. *Scientific American*, *120*(February), 5. http://doi.org/10.1126/scisignal.2009ec285

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2014) *Bayesian Data Analysis*, 3rd edition, Boca Raton, FL: CRC Press.

Hodgen, J., Anders, J., Bretscher, N. & Hardman, M. (In prepration). *SMART Spaces: Spaced Learning (Teaching Pilot Trial) Evaluation Protocol*. London: Education Endowment Foundation.

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

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*(1), 178-206. doi:10.3758/s13423-016-1221-4

Nieuwenhuis, S., Elzinga, B. M., Ras, P. H., Berends, F., Duijs, P., Samara, Z., & Slagter, H. A. (2013). Bilateral saccadic eye movements and tactile stimulation, but not auditory stimulation, enhance memory retrieval. *Brain and Cognition*, *81*(1), 52–56. http://doi.org/10.1016/j.bandc.2012.10.003

Nitz, S., Prechtl, H., & Nerdel, C. (2014). Survey of classroom use of representations: development, field test and multilevel analysis. *Learning Environments Research, 17*(3), 401-422. doi:10.1007/s10984-014-9166-x.

O'Hare, L., Stark, P., McGuinness, C., Biggart, A., & Thurston, A. (2017). *Spaced Learning: The Design, Feasibility and Optimisation of SMART Spaces: Evaluation report and executive summary*. London: Education Endowment Foundation.

Royal Society. (2014). *Vision for science and mathematics education*. London: The Royal Society.

Smallwood, N., Spriggs, M. J., Thompson, C. S., Wu, C. C., Hamm, J. P., Moreau, D., & Kirk, I. J. (2015). Influence of Physical Activity on Human Sensory Long-Term Potentiation. *Journal of the International Neuropsychological Society*, *21*(10), 831–840. http://doi.org/10.1017/S1355617715001095

Smolen, P., Zhang, Y., & Byrne, J. H. (2016). The right time to learn: mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience, 17*, 77. doi:10.1038/nrn.2015.18

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*(6), 837-848. doi:doi:10.1002/acp.1598

TISME. (2013). *What influences participation in science and mathematics? A briefing paper from the Targeted Initiative on Science and Mathematics Education (TISME)*. London: TISME / King's College London.

Wake, G., Hodgen, J., & Adkins, M. (2017). *Young Enterprise: Mathematics in Context Evaluation Protocol.* London: Education Endowment Foundation.

Winter, B., Breitenstein, C., Mooren, F. C., Voelker, K., Fobker, M., Lechtermann, A., Krueger, K., Fromme, A., Korsukewitz, C., Floel, A., & Knecht, S. (2007). High impact running improves learning. *Neurobiology of Learning and Memory, 87*(4), 597-609. doi:https://doi.org/10.1016/j.nlm.2006.11.003

Appendix 1: Logic model and data collection

Figures 4a and 4b indicate how data collected link to key elements within the logic model.

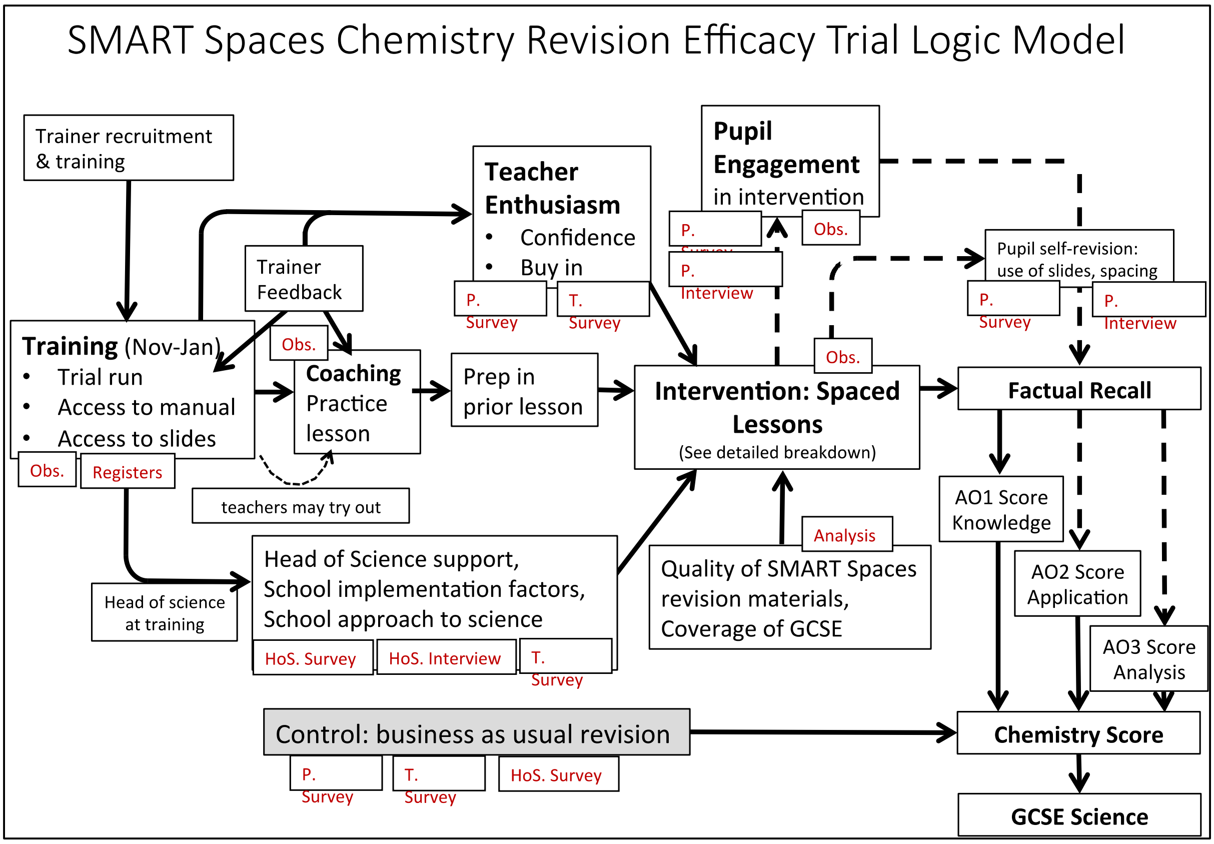

Figure 4a: SMART Spaces Chemistry Revision Logic Model (overall), with IPE measures

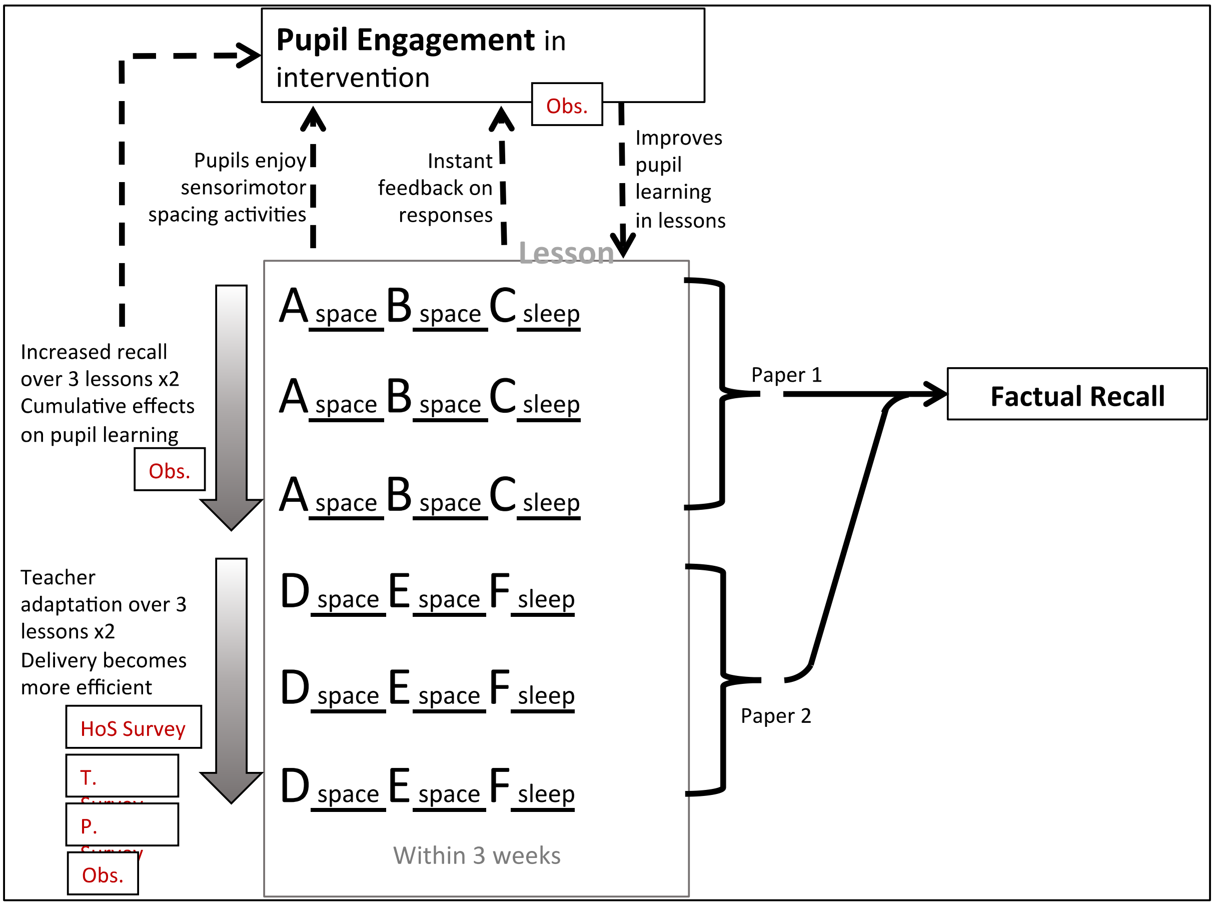

Figure 4b: SMART Spaces Chemistry Revision Logic Model (intervention spaced lessons element), with IPE measures

Appendix 2: Statement of legal basis for processing data

As part of this project, we process pupils' personal data. For this reason, it is important that we process this data lawfully, following the principles laid out in the Data Protection Act 1998 (DPA) until May 2018 and the General Data Protection Regulation (GDPR) thereafter. We explain the lawful basis below with respect to the GDPR but there are equivalent regulations in the DPA for the justifications set out below.

We use Article 6(1)e of the GDPR as the lawful basis for processing personal data as part of this project. This is generally known as the "public task" basis. UCL has reviewed current ICO guidance available here: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/, and has determined that this research forms part of its performance of a task in the public interest, as one of its core purposes provided for in its Charter and Statutes. We do not believe that any of the data we process falls within the definition of special category data under the GDPR. This would require an additional justification under Article 9(2) of the GDPR.

In order to use the public task basis we set out below how this is a task in the public interest and demonstrate that the processing is necessary to achieve the purpose of the processing.

Public benefit: Use of pupil's personal data as part of this evaluation is to understand the benefits to pupils, teachers and schools of participating in the SMART Spaces programmes in chemistry education in terms of academic attainment, improved pedagogy and other related benefits. This has public benefits that we believe are significant in terms of understanding whether this programme has the potential to benefit children in schools across England. If we could not do this then it would not be possible to provide this new evidence. Our proposed research has been reviewed by the UCL Institute of Education research ethics committee [REC1052] and the UCL Data Protection team [Z6364106/2018/03/25 social research], meaning we believe our use of the data to be ethical and lawful.

Necessity: This processing does help to further the interest of providing evidence on what works in
promoting academic attainment among pupils in English schools by providing high-quality evidence based on a sufficiently robust design. For the evaluation of the SMART Spaces Revision version, we do this using a randomised controlled trial (RCT) together with a mixed-methods implementation and process evaluation (IPE) to gather evidence about *inter alia* the necessary conditions for success. This is a recognised high-quality research design applied internationally to provide evidence of this type, meaning we consider this is a reasonable approach. For the evaluation of the SMART Spaces Teaching version, we propose a pilot study that will collect evidence of the promise, feasibility and scalability of the intervention, which we consider to be a reasonable approach. It would not be practical in either case to provide this quality of evidence without processing pupils' and teachers' data.

NPD Access: When applying for NPD data the relevant lawful reason for requesting that data will be that our task is specified in the Education (Individual Pupil Information) (Prescribed Persons) (England) Regulations 2009: Regulation 3 (1)(b) and (6)(d), including as amended by the Education (Individual Pupil Information) (Prescribed Persons) (England) (Amendment) Regulations 2013.