Education Endowment Foundation

SMART Spaces: Spaced Learning Revision Programme

Evaluation Report

July 2023

Jeremy Hodgen, Nicola Bretscher, Mark Hardman, Jake Anders, Helen Lawson

UCL IOE – FACULTY OF EDUCATION AND SOCIETY

The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries and colleges to improve teaching and learning for 2 – 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

For more information about the EEF or this report please contact:

Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from the Institute of Education at University College London's Faculty of Education and Society: Jeremy Hodgen, Nicola Bretscher, Mark Hardman, Jake Anders, and Helen Lawson.

The lead evaluator was Jeremy Hodgen.

Contact details

Prof Jeremy Hodgen
UCL Institute of Education
Bedford Way, London WC1H 0AL
jeremy.hodgen@ucl.ac.uk

Acknowledgements

The research team and authors of this report would like to thank the many teachers and their pupils who took part in the trial—both those who took part in the SMART Spaces intervention programme and those in the control group. The time and effort they gave to ensure that we had adequate data to work with was essential and we are very grateful to them.

We would also like to thank the developers at Queen's University Belfast and Hallam Teaching Schools Alliance who worked on setting up and ensuring the smooth running of the SMART Spaces intervention: their work and support has been invaluable at every stage. We are especially grateful to Liam O'Hare, Alastair Gittner, John Coats, Maria Cockerill, Patrick Stark, and Aideen Gildea.

The team at the EEF have also been extremely helpful, particularly Danni Mason and Diotima Rapp.

We are also grateful to AQA and its staff for their help with accessing pupil's item scores, particularly Julian Clarke, AQA Head of Curriculum – Science.

Finally, we are grateful to colleagues at UCL who worked on the project, particularly Katie Pepper, Haira Gandolfi, and Bohan Lui.

Executive summary

The project

SMART Spaces: Spaced Learning Revision Programme (SMART Spaces Revision) aims to raise attainment in GCSE chemistry by improving revision using ‘spaced learning’. Spaced learning involves teachers repeatedly delivering the same content, across multiple sessions, with breaks in between.

The programme, developed and delivered by Hallam Teaching School Alliance (HTSA) and Queen’s University Belfast (QUB), trains teachers to deliver six highly structured and manualised chemistry revision lessons for the AQA Combined Science GCSE. Lessons are provided to Year 11 pupils (age 15 to 16) in the three-week period prior to GCSE examinations. Each lesson consists of three 12-minute episodes delivering content separated by two ten-minute spacing activities such as juggling. Lessons are delivered on three separate days, which allows additional spaces of around 24 hours between content repetition. The first three lessons repeatedly cover one half of the curriculum and the second three lessons repeatedly cover the other half.

SMART Spaces Revision training involves a lead teacher who supports implementation and chemistry teachers who receive a day of instruction (in twilight sessions). In-school follow-up support and teaching resources are also provided. Resources include PowerPoint slides covering the entire chemistry curriculum content for the AQA Combined Science award together with a manual, an activity pack, and spacing materials.

This project was a two-armed, cluster-randomised controlled efficacy trial (c-RCT) with 125 schools from across England: 54 schools were randomly allocated to receive the intervention and 71 acted at the ‘business as usual’ control group; 14,098 pupils from Year 11 taking the AQA Combined Science GCSE took part in the trial. The evaluation tested the impact of the intervention on GCSE chemistry attainment with surveys and interviews with teachers and pupils and observations of training informing the process evaluation. The evaluation was conducted by the IOE at UCL’s Faculty of Education and Society and delivery occurred between April and May 2019.

Table 1: Key conclusions

Key conclusions
1. Pupils in the SMART Spaces Revision schools, on average, made no additional months’ progress in chemistry attainment in AQA Combined Science GCSE compared to pupils in the control schools. This result had a high security rating.
2. Pupils in the SMART Spaces Revision schools, on average, made no additional months’ progress in overall GCSE Combined Award Science compared to pupils in the control schools.
3. Pupils eligible for free school meals (FSM) in the SMART Spaces Revision schools, on average, made no additional months’ progress in chemistry attainment at AQA Combined Science GCSE compared to FSM-eligible pupils in the control schools.
4. The SMART Spaces Revision theory of change predicts that spaced learning primarily improves recall of knowledge. There was weak evidence that intervention pupils had improved chemistry knowledge attainment compared to the control pupils. However, the effect was small and did not meet the threshold set by the evaluator that it was non-zero. Also, it may be at the expense of a negative impact on chemistry application and analysis attainment.
5. The SMART Spaces Revision intervention was delivered as intended in schools, indicated through observations, pupil and teacher surveys, and interviews. Teachers showed a high level of support for the intervention based on teacher surveys and interviews, with 70% reporting that they would be happy to deliver SMART Spaces Revision in the future.

EEF security rating

These findings have a high security rating. This was an efficacy trial designed to test whether SMART Spaces Revision worked under developer-led conditions in a number of schools. The trial was a well-designed, two-armed, cluster-randomised controlled trial. The trial was well-powered and pupils in the intervention were similar to those in the comparison schools in terms of prior attainment. However, 15.1% of the pupils who started the trial were not included the final analysis because of school withdrawal, which reduced the security of the trial.

Additional findings

Pupils in SMART Spaces Revision schools made, on average, zero additional months' progress in AQA Combined Science GCSE chemistry compared to those in the control group. This is our best estimate of impact, which has a high security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes negative effects of two months' less progress and positive effects of up to two months' additional progress.

Analysis was undertaken to explore whether certain pupil characteristics might influence the impact of the intervention. This included whether the impact was different for pupils in receipt of free school meals (FSM pupils), girls compared to boys, or pupils with low compared to high attainment. In each instance there was no evidence of differing impact.

AQA Combined Science GCSE chemistry has three areas of assessment: knowledge (AO1), application (AO2), and analysis (AO3). Exploratory analysis of pupil attainment in each of these assessment areas suggests that pupils in SMART Spaces schools made, on average, one additional months' progress in knowledge (AO1). This contrasts with the AO2 and AO3 assessment areas where pupils in the intervention made one months' less progress compared to pupils in the control schools. These results reflect the impact, on average, in each assessment area, however, there is a wide range of potential effect sizes for AO1 – AO3, which suggests that the possible impact of SMART Spaces Revision may vary greatly for each of these assessment areas.

The SMART Spaces Revision theory of change predicts that spaced learning is the mechanism through which factual recall is improved, and thus attainment. AO1 is focused of factual recall so is more closely aligned to the content and aims of SMART Spaces Revision. There is, therefore, some weak evidence to suggest that the SMART Spaces intervention improved recall in AO1, which spaced learning may have supported. However, these gains did not extend to chemistry attainment overall and, for AO2 and AO3, the gains were negative.

SMART Spaces Revision was well received by schools, teachers, and pupils, which is reflected in self-reported surveys. The intervention was delivered as intended by the developers in terms of delivery by schools, lesson aims and pedagogic approach. Teachers reported that the training, support, and resources supplied made the intervention easy to deliver.

The results indicate that this particular approach, SMART Spaces Revision, did not have an additional impact on pupils' overall chemistry attainment when compared to usual practice. The current trial provides some evidence that teachers are willing to implement such approaches but the challenge of translating the broad research findings about learning involving spacing into actionable teaching strategies and interventions remains.

This trial should not be interpreted as providing definitive evidence for or against the efficacy of spaced learning more generally, and the wider evidence base should also be considered. There remains a need for research focused on the design, implementation, and efficacy of spaced learning teaching approaches and interventions.

Cost

The average cost of SMART Spaces Revision for one school was around £3,270, or £9 per pupil per year, when averaged over three years. This expenditure falls in the first year, which includes training, in-school coaching, and resources. There are limited additional costs to schools as training occurred during twilight sessions.

Impact

Table 2: Summary of impact on primary outcome

Outcome/ group	Effect size (95% credible interval)	Estimated months' progress	EEF security rating	No. of pupils	Region of Practical Equivalence (ROPE)	EEF cost rating
Chemistry subscale of GCSE combined science	-0.01 (-0.17, 0.12)	0	4 of 5 padlocks	11976 (5309; 6667)	84%[1]	£ £ £ £ £
Chemistry subscale of GCSE combined science —FSM pupils	-0.06 (-0.14, 0.01)	-1	N/A	3142 (1466, 1676)	83%[1]	£ £ £ £ £

[1] The Region of Practical Equivalence (ROPE) is a Bayesian approach to assessing any difference from the null, or zero, value (Kruschke, 2018). ROPE indicates the proportion of the parameter (effect size) values that are equivalent to the null value for practical purposes. In these cases, 84% and 83% of the distribution of the effect size, respectively, is practically equivalent to zero.

Introduction

This evaluation report presents the results from an evaluation of SMART Spaces: Spaced Learning Revision Programme (SMART Spaces Revision), an efficacy trial funded by the Education Endowment Foundation (the 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 was carried out by a team from the the IOE at UCL’s Faculty of Education and Society (UCL)—’the evaluator’.

The evaluation consisted 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 took place over the 2018/2019 academic year, with randomisation taking place in two stages in October and December 2018 and delivery of the intervention in schools during April and May 2019 following teacher training and coaching sessions over the period December 2018 to April 2019.

This trial follows a promising evaluation of a pilot study, also funded by the EEF (O’Hare et al., 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 report (Hodgen et al., 2022).

Background

Spaced learning is a promising development for science education (with important implications for other subjects such as mathematics and languages). 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 (Ebbinghaus, 1913).

O’Hare et al.’s (2017) review of the literature on the spacing effect indicated that there is robust evidence from experimental studies in the laboratory to support a positive effect of spaced learning strategy on memory and retention (when compared to massed learning). For example, Cepeda et al. (2006) conducted a quantitative synthesis of 317 experiments and found that all but 12 of the studies indicated a positive benefit of spaced learning on recall accuracy (see also, Dononvan and Radosevich, 1999). There is also some evidence to suggest that spaced learning may be effective for complex tasks as well as simple tasks (see, for example, Miles, 2014). However, the mechanisms underlying the spacing effect are poorly understood and there are several competing theories of how spacing affects learning (see, for example, Smolen, Zhang and Byrne, 2016). Moreover, much of the research was conducted in laboratory settings and translating this body of evidence into effective classroom strategies is less well understood.

A significant question relates to the optimal length and organisation of spacing. Short spaces (ten minutes) have been found to be effective in the neuroscience literature (Stern, 2009) although less is known about the role a particular activity could play within a short space for improving learning. O’Hare et al. (2017) also examined the evidence about what happens *during* the space and concluded that sleep is important for memory formation (Bell et al., 2014). The original pilot study addressed this question by conducting a feasibility study (FS) and an optimisation study (OS) examining the effects of short spaces (ten-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 in addition to spacing, and O’Hare et al. (2017, p.35) note that a nuanced understanding of the effects of this (Taylor and Rohrer, 2010) has yet to be explored in relation to SMART Spaces Revision. The most promising approach in the OS was the combination of ten-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.

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 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. The developers hypothesised that, since both of these potential facilitators of memory formation are present during juggling (saccadic eye movements when tracking balls and alternating tactile stimulation when catching), that juggling during spaces between learning would facilitate memory formation (O'Hare et al., 2017). 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 only investigated in the original pilot. 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.

While the evidence for the effect is robust within psychological testing, the integration of this within teaching practice in ordinary school classrooms is an area where there is not yet a strong evidence base. In their systematic review, Perry et al. (2021) identified only three studies judged to have high potential as a basis for assessing the effectiveness of spacing across lessons—one of which was O'Hare et al.'s (2017) pilot study of SMART Spaces Revision. Based on these and a number of weaker studies, Perry et al. concluded that spaced practice may have a small positive effect on attainment in science and mathematics when compared to non-spaced, or massed, practice but also concluded that their confidence in the effect was limited. This systematic review indicates a need for further high quality research into the effect of spacing and suggests that science is likely to be a productive subject for this intervention. Moreover, the earlier pilot of SMART Spaces for Revision provided evidence of the feasibility of the specific SMART Spaces approach to spacing across lessons and its readiness for implementation in a large efficacy trial. Additionally, a small randomised controlled trial (RCT) within the earlier pilot provided evidence of the promise of the SMART Spaces Revision intervention to result in improved performance on GCSE chemistry items. This RCT tested different models of spacing between condensed content delivery and provided some preliminary evidence that the most promising approach to spaced learning combined the use of both ten-minute and 24-hour spaces.

Intervention

The intervention is a further development of the original pilot (O'Hare et al., 2017), designed to be delivered at scale across the 54 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 (see Figure 3 and Figure 4). A SMART Spaces manual provides further guidance for teachers and schools.

1. Brief name

SMART Spaces Revision Version.

2. Why—rationale/theory

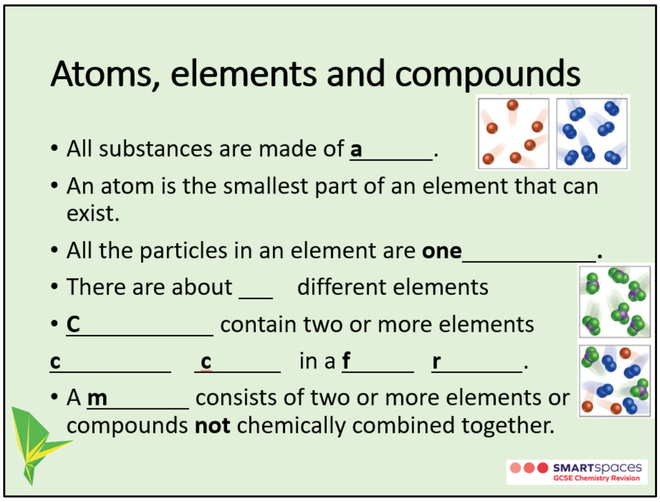
SMART Spaces Revision is an educational programme for AQA GCSE combined award science pupils used for examination revision in chemistry using spaced learning. Evidence from the neuroscience and cognitive psychology indicates that including spaced intervals between learning sessions can improve factual recall, and an earlier pilot suggested that a combination of short (ten-minute) and longer (approximately 24-hour or period of sleep) spaces provides a promising model of spacing (O'Hare et al., 2017). Therefore, it is the spacing that is of fundamental importance in this intervention. It was anticipated that factual recall would 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 (15 and 16 years old) in schools across England entered for GCSE combined award science.

4. What—materials

PowerPoint chemistry revision slides covering the entire GCSE double science chemistry curriculum are used in intervention spaced lessons. There is one set of slides covering content for AQA GCSE chemistry Paper 1 and a further set for Paper 2. In the second and third run-through, the slides were adapted to include missing words or labels for the pupils to recall aloud. See Figure 1 and Figure 2: Example slides for AQA Combined Science Trilogy chemistry Paper 1—second run through



for examples of these slides. Materials also include a SMART Spaces manual and SMART Spaces activity pack to be used by teachers and materials for spacing activities during intervention spaced lessons (for example, juggling balls).

Figure 1: Example slides for AQA Combined Science Trilogy chemistry Paper 1—first run-through

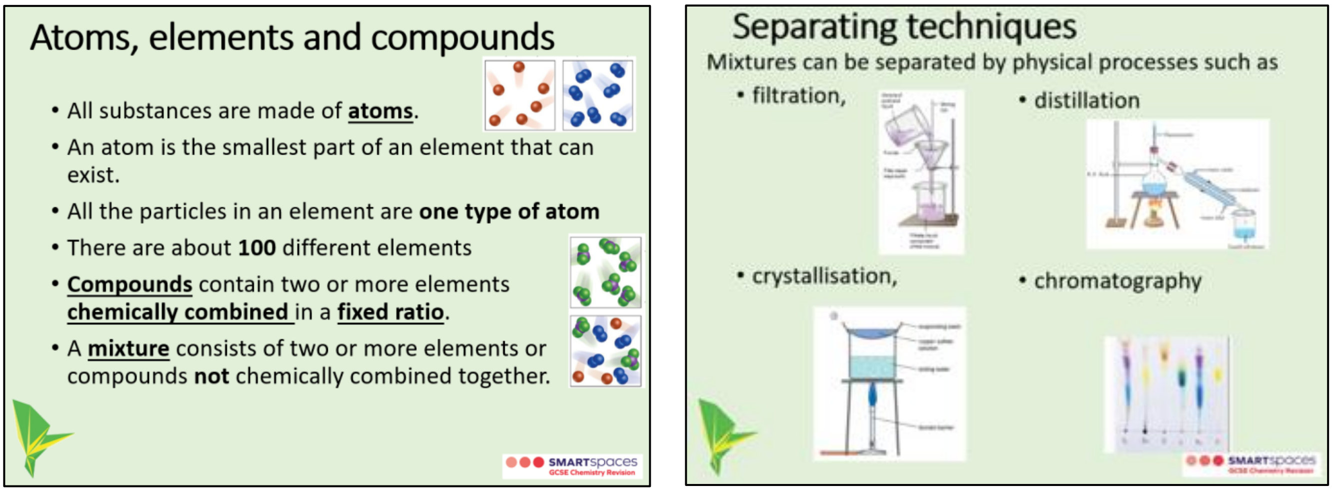
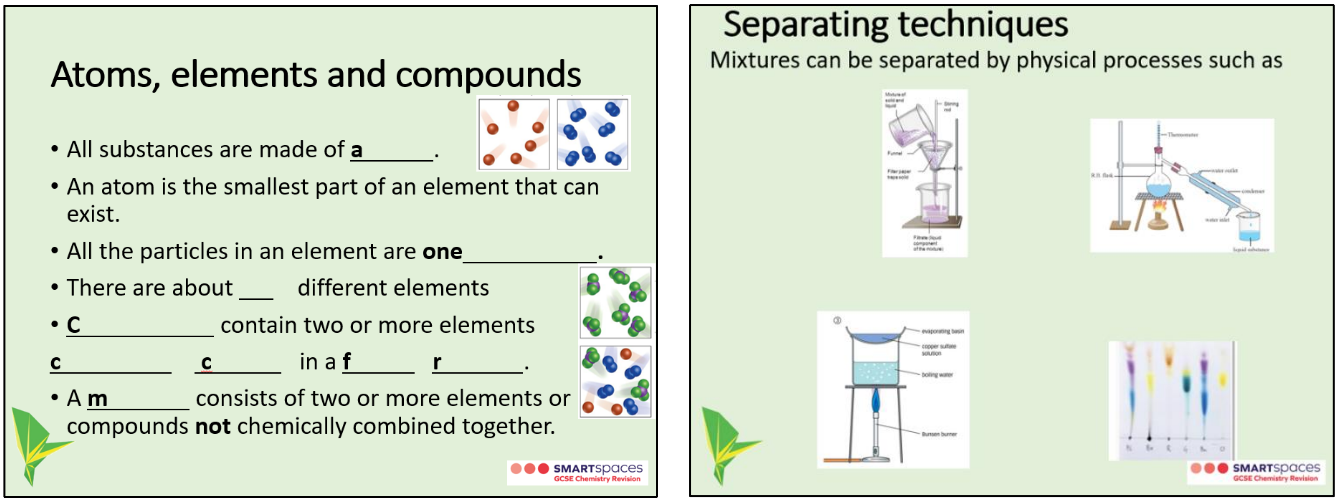


Figure 2: Example slides for AQA Combined Science Trilogy chemistry Paper 1—second run through



5. What—procedures

The intervention consists of two elements: SMART training and SMART Spaces implementation. The training was delivered by one of the developers (with extensive experience as a chemistry education specialist) and by doctoral students and post-doctoral researchers from QUB who themselves were trained by the experienced developer.

SMART training

All teachers who were to deliver SMART Spaces Revision and the head of department (or designated SMART Spaces lead teacher) were trained in a half-day training session, which included 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 received 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 took place between January and April 2019. During coaching visits, each teacher was to be observed delivering at least a 15- to 20-minute segment of a SMART Spaces lesson, including teacher delivery and transition to and from a space. It was expected that these would likely be with examination classes (affecting dosage—some pupils received more than the prescribed number of iterations of the material). In practice, these were sometimes the first delivery sessions of the SMART Spaces Revision intervention. Trainers complete a standard pro forma to provide feedback to teachers who have been trained. Feedback was also provided either face to face or by phone and covered timing, content delivery, evidence of teacher preparation, classroom organisation, and transitions.

Teachers were encouraged, but not required, to further practice spacing activities with different classes prior to the intervention (affecting dosage for teachers). Some teachers also practiced without their classes or with colleagues.

SMART Spaces implementation

Pupils were prepared for the intervention in a prior lesson by the teacher explaining that the following revision lessons would use the SMART approach.

Chemistry topics for AQA Paper 1 were to be taught using the PowerPoint slides in the three short approximately 12-minute sessions, A, B and C, with ten-minute spaces between each topic: A-space-B-space-C. Additional spacing is assumed to occur before and after the lesson (-A and C-) due to changes in activity. Where SMART Spaces lessons took place in the second part of a double lesson, developers suggested that 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 provided schools with juggling balls and advised them that juggling was used in the pilot study and found to be feasible in the classroom. They also provided a list of alternative activities and advised schools they may choose from this list if they do not feel juggling is practical within their lessons. These were:

- balloon games—for example, two lines of pupils racing to pass balloon from front to back;
- modelling clay;
- origami; and
- ‘Simon says’.

This full lesson (A-B-C) was to be 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. Who—implementers

The SMART Spaces Revision intervention lessons were delivered by GCSE science teachers who have had SMART training. The same teacher provided the whole SMART Spaces Revision programme, ideally in two consecutive weeks within the three-week intervention window.

The SMART training and coaching was provided by trainers experienced in the delivery of SMART Spaces Revision. Heads of science (or designated SMART Spaces lead teachers) were also required to be present at the training to ensure that departmental implementation was coordinated and supported and that the SMART Spaces Revision intervention was integrated within the school's overall approach to science revision.

7. How—mode of delivery

A whole-class programme that was conducted during six normal science lessons. SMART training was delivered to groups of teachers. SMART coaching was delivered to teachers by trainers in one to one sessions following observations of a practice SMART session.

8. Where—setting

SMART training was conducted in out-of-school (or twilight in-school) sessions, SMART coaching took place in school, and the SMART Spaces Revision intervention itself in standard GCSE classrooms.

9. When and how much—dosage

The programme of six intervention spaced lessons covers the AQA GCSE chemistry curriculum in a high intensity way and was delivered in two blocks of three days during the three-week period prior to the first GCSE double award examination. The SMART Spaces Revision slides are structured in six 12-minute chunks of GCSE chemistry content (approximately one sixth of each GCSE 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 was an expectation that a teacher's delivery of the 12-minute chunks would become more efficient over the three consecutive days as less repetitive elaboration would take place. Pupils experienced six one-hour periods in total.

For this trial, 23 April to 10 May 2019, inclusive, was the window in which the intervention groups were required to undergo three iterations of Paper 1 content (A-B-C) followed by three iterations of Paper 2 content (D-E-F). This was organised around the exam timetable for summer 2019, which indicates that chemistry Paper 1 is scheduled for 16 May and chemistry Paper 2 for 12 June 2019.

10.Tailoring

SMART Spaces Revision is a manualised intervention and treatment fidelity should be maximised. Teachers could choose from a menu of spacing activities. It was expected that teachers would become more efficient over the three iterations of A-B-C, and likewise the three iterations of D-E-F. This is due to teacher learning, less verbal embellishment by the teacher, and better recall by pupils. This was theorised as allowing for some adaptation of the time spent on particular topics and the provision of more feedback to pupils.

The intervention materials included slides which covered specification content relevant only to higher tier students. These were marked 'higher tier' and teachers were told that they could skip these slides if their class was working to foundation tier. As a consequence, classes covering higher tier materials had more to cover within the SMART Spaces lessons. Otherwise, however, there was no difference to the delivery of materials or coverage of specification topics between higher and foundation tiers.

The expectation was that some teachers may share slides with pupils (although this will not be actively encouraged). Pupils engaged in the intervention may therefore use the slides or adopt spacing within their own revision practices.

11.How well planned

Effective implementation required training for all teachers who taught Year 11 GCSE combined award chemistry in all trial schools before they delivered the intervention spaced lessons. This training consisted of modelling, practice, and feedback on programme delivery. It was anticipated that teacher enthusiasm would influence the fidelity and quality of delivery of the intervention spaced lessons as well as whether those lessons were delivered. Effective implementation also required support from a head of science (or designated SMART Space lead teacher) to ensure that lessons were scheduled and supported within the revision period.

Figure 3: SMART SPACES Revision logic model—overall

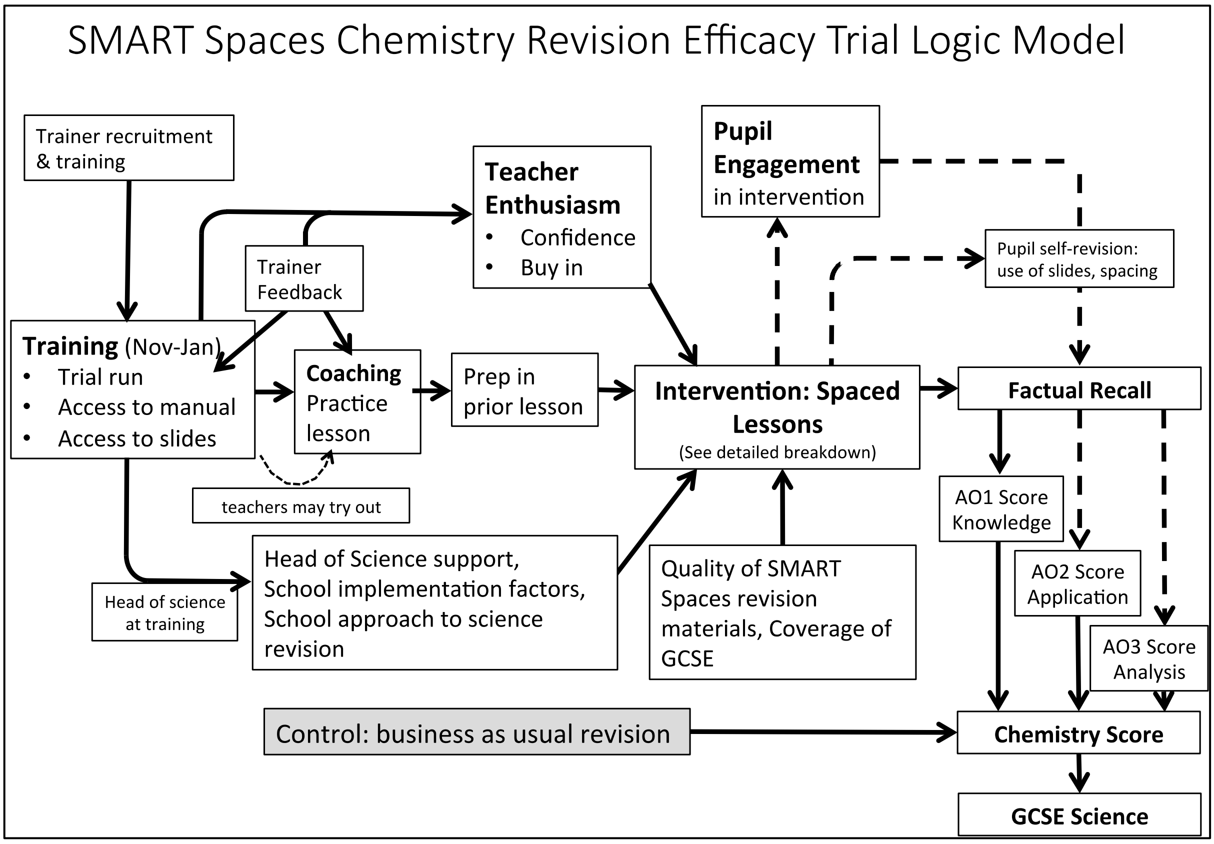
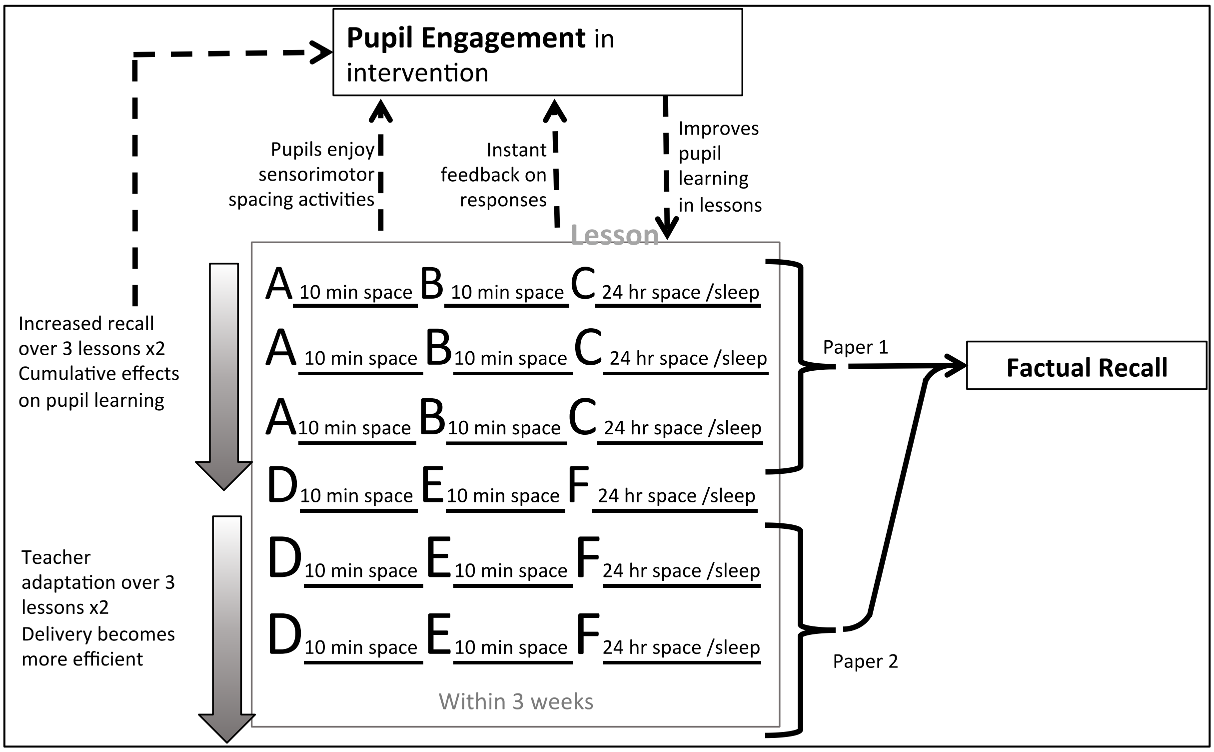


Figure 4: SMART Spaces Revision logic model—intervention spaced lessons element



Evaluation objectives

The evaluation design and methods were pre-specified in an evaluation protocol and a statistical analysis plan (SAP), which are available on the EEF website.[2]

The trial addressed the following primary research question.

RQ1. What is the size of the effect of the SMART Spaces Revision 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 addressed the following secondary research questions.

RQ2. What is the size of the effect of the SMART Spaces Revision 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 Revision intervention on pupils' attainment in the assessment objectives (AOs) 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 different assessment objectives, and are the effects practically distinguishable from a null effect?

RQ4. What is the size of the effect of the SMART Spaces Revision 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?

In the process evaluation, we addressed the following research questions. 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 Revision implemented with fidelity in the trial, and to what extent can SMART Spaces Revision 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 Revision with fidelity? [Quality, (teacher) responsiveness, implementation.]

RQ10. Do teachers and heads of science perceive SMART Spaces Revision 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 Revision? [Adaptation.]

[2] https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/smart-spaces

RQ14. (a) Are all pupils responsive to SMART Spaces Revision 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 Revision distinguishable from 'business as usual' revision practice in schools? [Differentiation, monitoring of control group, implementation.]

RQ17. To what extent does the logic model (Figure 3 and Figure 4) adequately describe the mechanism by which the SMART Spaces Revision intervention effected change (if any)? [Differentiation, intervention characteristics.]

Ethics and trial 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 April 2018 by SSESW, QUB Research Ethics Committee. The trial was pre-registered on the ISRCTN registry, reference number 54008927.[3]

The headteachers of schools wishing to take part in the trial were required to sign a memorandum of understanding (see Appendix C).

We processed personal data for public interest purposes (see Data Protection section below). Nevertheless, we provided an opportunity for parents and 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 had no objection to their data being processed in this way. This demonstrated that the processing does not impinge on anyone's rights and meets our responsibilities under the BERA Ethical Guidelines for Educational Research (particularly regarding informed consent, openness, and disclosure).[4]

Parents, and participating pupils, were 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 explained the intervention and the research being conducted in simple language, provided opportunities for parents to ask additional questions, and provided clear steps to follow if they wished their child to be withdrawn from any data processing as part of the research. The sheet and form also made it clear that data could be withdrawn at this point or at any point up to 31 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) involved more active participation of teachers and pupils, including lesson observations and interviews. To this end, we collected 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. Information sheets and consent forms are included in Appendix D.

Data protection

Data has been 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). UCL and QUB are both data controllers and each has processed data on the basis of the 'public task' purpose (as per condition 6(1)e of the GDPR). HTSA is a data processor and processed data on the basis of the 'legitimate interest' purpose (as per condition 6(1)f of the GDPR). UCL has reviewed current ICO guidance[5] and has determined that this research forms part of its 'performance of a task in the public interest'—one of its core purposes according to its Charter and Statutes.

---

[3] https://doi.org/10.1186/ISRCTN54008927
[4] https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018
[5] https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/

We do not believe that any of the data we processed 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, were informed of the proposed data processing and given an opportunity to object to this and withdraw their, or their child's, data. The information provided to parents and carers, pupils, and teachers explained in clear and plain non-technical language the purpose to which we put the data, that they could object to this data usage, and that this would be respected. They were also informed of the contact details of the organisation and categories of data that we processed and that the data processing has been compliant with the GDPR and data protection legislation.

The evaluation team at UCL carried out a data protection impact assessment and put in place a data management plan. As part of this plan, data was checked and cleaned to ensure the GDPR principle (d) of accuracy is met. See Appendix E for the privacy notice, statement of the lawful basis, and public tasks assessment for data processing.

Data security

All personal data collected or obtained as part of this project has been treated as 'highly restricted' under UCL Data Protection classification guidance. Personal data was collected from schools (pupil names, UPNs, dates of birth, FSM eligibility, sex, national test results, class, and teacher, as well as teacher names and survey data). In addition, with permission from the schools, item-by-item mark data for the Combined Science GCSE were provided directly by the examination board, AQA, under a data sharing agreement with UCL. This data was stored, processed, and analysed on the UCL Data Safe Haven (DSH), the technical infrastructure that UCL has built specifically to host sensitive research data. This data was matched to data from the National Pupil Database (NPD) by the Department for Education (DfE), NPD licence DR190621.01, and placed in a pseudonymised format in the Office for National Statistics Secure Research Service (ONS SRS) where the impact analysis was conducted.

Qualitative data was also pseudonymised. Once pseudonymised it was 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 recordings are 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 was required between collaborators on this project at UCL and QUB, and between UCL and AQA. This will be conducted by making a secure remote connection (for example, VPN) between the university networks and transferring data across this. In addition, the data was encrypted before sharing using a password shared between research team members by separate communication.

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

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

A risk assessment was conducted for the storage, processing, and transfer of all personal data for the SMART Spaces project. All team members undertook regular annual data security training and all those accessing the ONS SRS were ONS accredited researchers.

The DSH environment is certified to ISO27001:2013 with BSI, certificate number: IS 612909. The most recent external audit, prior to the commencement of the trial, was in May 2017. The hosting was 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 ongoing 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. It will be kept for no longer than ten years in line with UCL's guidance on retention of records for research.

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

Project team

QUB and HTSA 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 co-designer and 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.

Aideen Gildea (QUB): school support, assisted with data collection.

Dr Joanne O'Keefe: school support.

Ewan MacRae (QUB): PhD researcher, teacher CPD.

UCL Evaluation Team

Professor Jeremy Hodgen: project lead, overall direction and impact evaluation lead.

Dr Jake Anders: advice on the impact evaluation and statistical techniques.

Dr Nicola Bretscher: statistical analysis.

Dr Mark Hardman: leading the IPE.

Dr Helen Lawson: IPE fieldwork and analysis and contact with schools.

Dr Haira Gandolfi: IPE fieldwork and analysis.

Katie Pepper: administration and day to day project support.

Methods

Trial design

Table 3: Trial design

Trial design, including number of arms		Cluster randomised, two arms
Unit of randomisation		School
Stratification variable(s) (if applicable)		Randomisation block, school-level prior attainment
Primary outcome	Variable	Chemistry attainment
	Measure (instrument, scale, source)	Chemistry subscale of AQA GCSE Combined ('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, source)	(1) AQA GCSE Combined ('Double Award') Science (item-level, continuous) (2) Knowledge, application, and analysis of assessment objective (AO) subscales for the chemistry element of AQA GCSE Combined ('Double Award') Science (item-level, continuous)
Baseline for primary and secondary outcomes	Variable	Attainment at KS2 (combined English and mathematics scores)
	Measure (instrument, scale, source)	Weighted mean of marks in English (reading) and mathematics KS2 tests (KS2_READMRK; KS2_MATTOTMRK, 0-200, NPD)

The trial was a school-level randomised controlled trial in schools offering combined (or 'double award') science through the AQA board. It was a two-arm efficacy trial: SMART Spaces Revision compared to a business as usual control with chemistry score in the AQA GCSE Combined Science Award as the primary outcome and, as secondary outcomes, overall science scores and the knowledge, application and analysis assessment objectives (AO) subscales for the chemistry element of the AQA GCSE Combined Science Award.

No changes were made to the trial design as pre-specified in the evaluation protocol and SAP.[6] However, due to problems accessing software within the ONS SRS, the robustness check of carrying out a frequentist analysis was conducted only using the Ime4 package in R, and was not repeated using the MLWin software.

Participant selection

The trial involved approximately 14,000 Year 11 pupils from 125 schools. The developer had limited capacity to deliver the training and coaching to schools so, in order to ensure the trial had sufficient power given this limited capacity,

[6] https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/smart-spaces-trial-and-teaching-pilot

allocation to the arms was unequal: there were 54 intervention and 71 control schools. Control schools were offered a financial incentive of £1,000 following the completion of all evaluation requirements with staff and the required pupils in 2018 and 2019. Control schools were permitted to purchase the SMART Spaces Revision programme from QUB/HTSA for use from January 2020 (that is, after the trial had concluded).

The eligibility criteria for schools to participate were:

- participating schools must be state-funded secondary schools in England and have some of their pupils enrolled in AQA GCSE double award science;
- schools had to agree not to participate in another EEF GCSE science randomised trial that would interfere with implementation of the intervention with Year 11 pupils during 2018/2019 academic year; and
- schools had to return a signed memorandum of understanding (MoU) in which they committed to participating fully in the study, including the collection of outcome measures in summer 2019, regardless of the trial arm to which they are assigned.

Recruitment was led by the developer. Recruitment focused initially on three regions in England in order to facilitate training: the North and the East and West Midlands. Subsequently, recruitment was expanded to cover the whole of England. Recruitment aimed for the overall proportion of FSM pupils at school-level to be as high as, or higher than, the average across England.

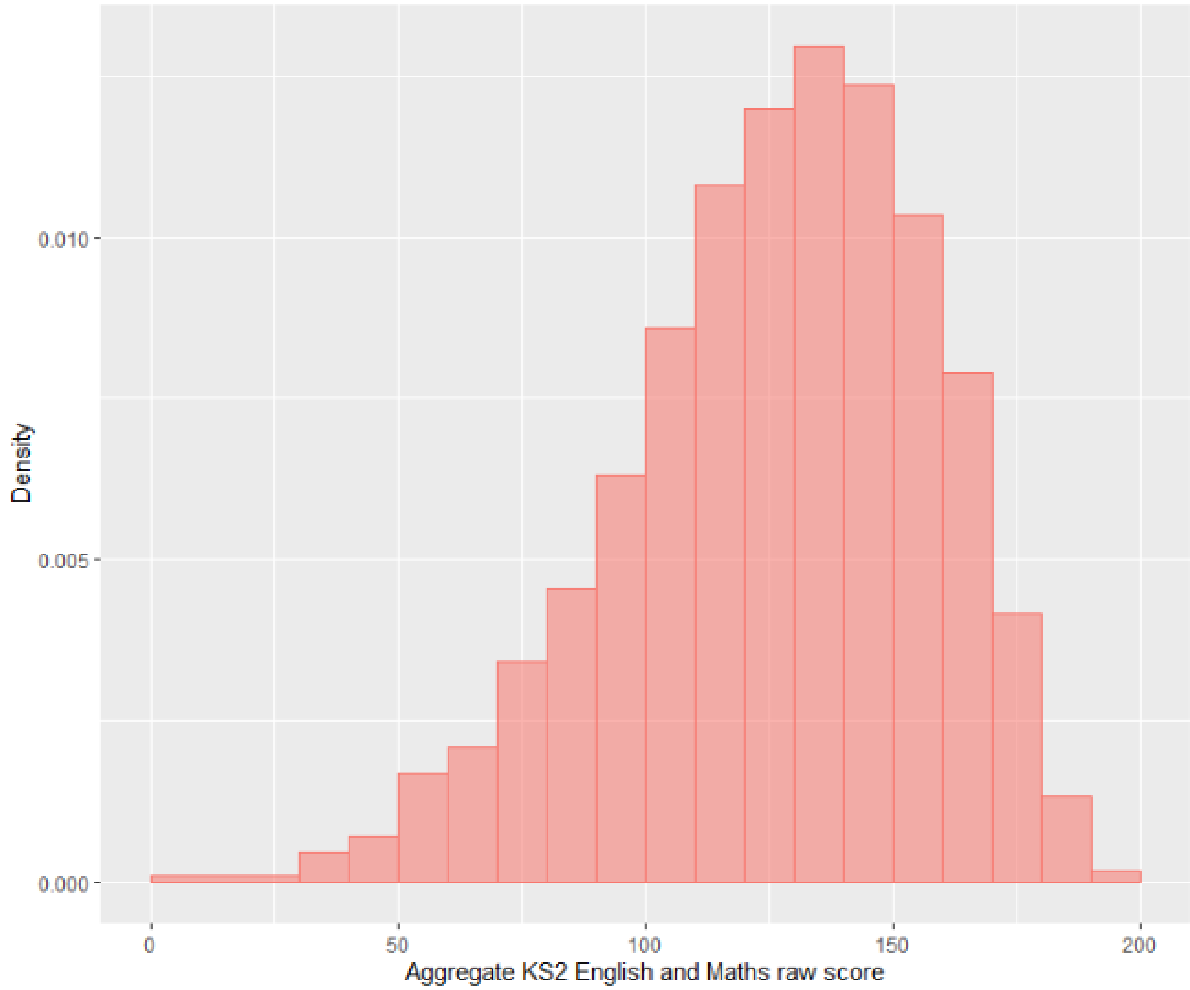
Outcome measures

Baseline measures

The key baseline measure for the primary analysis, and all secondary analyses using GCSE related outcomes, was the simple aggregation of KS2 National Test Scores in English and mathematics taken by all participating pupils at age 11 (the KS2 English and mathematics marks) in summer 2014. This was drawn from the DfE's National Pupil Database (NPD). As such, all participating schools were requested to provide pupil data (Unique Pupil Number, forename, surname, date of birth, gender, free school meals status, and school records of KS2 scores) in order to reliably carry out the matching to existing administrative data.

Figure 5 presents the distribution of the aggregated KS2 score for all participating pupils. We used a weighted aggregation to balance English (0–50) with maths (0–100) to get a score ranging from 0–200. As is typical for KS2 National Test Scores, there was some evidence of a ceiling effect. However, the measure had good statistical properties and correlations with the primary outcome measure were stronger than predicted for the power calculations in the evaluation protocol (see Table 9).

Figure 5: Distribution of aggregated KS2 National Test Score in English and mathematics



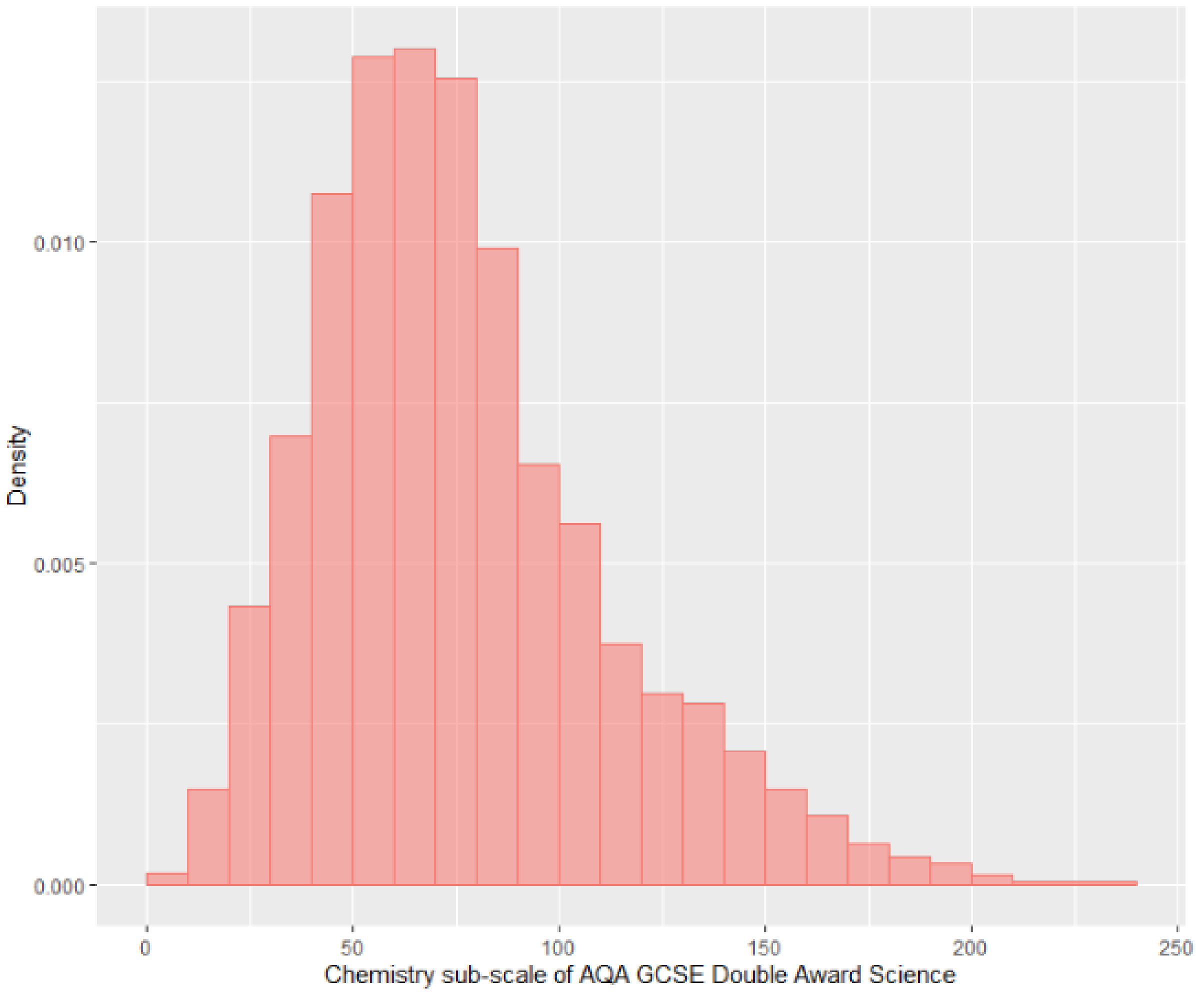
Primary outcome

The primary outcome is chemistry attainment at GCSE. This was measured using a chemistry subscale of AQA GCSE Combined Science based on item-by-item mark data for the 2019 examinations. This was 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.

AQA has two Combined Science specifications, Synergy and Trilogy, and each has two tiers of award, Foundation and Higher. We had originally planned to combine the scores for these different awards using the Uniform Mark Scale (UMS). Unfortunately, during the trial the exam boards ceased publishing UMS scores; instead, the total item-by-item marks for chemistry were equated using a linear scale generated from a least-squares regression on the three pairs of equated component grade boundary scores. See Appendix F for further details of the equating process. Item-by-item mark data was provided directly by AQA.

Figure 6 presents the distribution of the chemistry subscale. The distribution shows a positive skew in contrast to the negatively skewed pre-test measure. However, as noted earlier, the primary outcome measure has good statistical properties and correlations with the pre-test were stronger than predicted for the power calculations in the evaluation protocol (see Table 9).

Figure 6: Distribution of the primary outcome, the chemistry subscale



Secondary outcomes

There were four secondary outcomes:

- science attainment at GCSE, using a continuous numerical variable based on total scores; and
- the three AO subscales for the chemistry element of AQA GCSE Double Award Science (knowledge, application, and analysis), continuous numerical variables based on item-by-item mark data.

These measures were constructed on a similar basis to the primary outcome variable. See Appendix G for further details on the distributions of these measures.

Sample size

Sample size calculations were carried out with the aim of achieving a minimum detectable effect size, d, of 0.20 for the chemistry attainment measure since this was the primary outcome of interest. The results of MDES calculations at protocol, randomisation, and analysis stages of the trial are presented in Table 9 and discussed in the Participant Flow section below. These estimates were calculated using the PowerUpR package (Bulus et al., 2021) in R.

Protocol MDES calculations were based on the following assumptions:

- Randomisation would be at school level, stratified using a four blocked design. A number of options for unequal allocation to the intervention and control groups were considered before settling on the ratio 50:75 intervention to control. See Appendix H for the randomisation syntax.

- Number of children per cluster is 100. This is equivalent to around four double science GCSE classes per school. Since around 75% of pupils take the double award and the average size of a secondary school is around 180 pupils per year group, this is a relatively conservative assumption.
- A pupil-level pre- to post-test correlation of 0.5. No data is available for the correlation between the chemistry element of the GCSE double award science and combined Key Stage 2 (KS2) scores. The correlation was estimated on the basis of the correlation between GCSE science and combined KS2 scores being 0.556 (Benton and Sutch, 2014). This was judged a better predictor than the correlation between triple award chemistry and KS2 scores. However, since this correlation is lower (0.427, ibid.), a conservative estimate was judged appropriate.
- School-level pre- to post-test correlation of 0.25. This was estimated to be half the pupil-level correlation on the basis of advice from the EEF evaluation advisory panel.
- An intra-cluster correlation coefficient (ICC) of 0.15, based on the EEF's (2015) guidance.
- Power: 80%; significance level: 5%. These are standard assumptions.

Randomisation MDES calculations were based on the same assumptions, with the following alterations:

- Randomisation took place in two batches, due to recruitment difficulties, with schools stratified into 24 blocks on the basis of school-level KS2 prior attainment in a form of pair matching. Schools were allocated in a ratio 54:71 intervention to control.
- Number of children per cluster is 113. This was calculated as the mean number of GCSE double award science pupils for whom data was submitted per school.
- Number of pupils eligible for FSM per cluster is 25. This was calculated as the mean number of pupils—out of those included in the school sample—identified as eligible for FSM per school.

Statistical analysis

Primary analysis

Our primary analysis focused on the chemistry subscale of the AQA GCSE Combined Science award as described above. Outcome variables were modelled on the basis of intention to treat (ITT) using a multilevel linear model. We fitted a two-level model of pupils clustered in schools incorporating the treatment condition, the pre-test, and other stratification variables used for randomisation as covariates. These stratification variables were the block assignment based on KS2 school-level average attainment and whether the school was randomised as part of the first or second batch. Note that including pupil-level KS2 attainment as a covariate in the analysis model is the appropriate adjustment for both our re-randomisation and our blocking strategies (Rubin, 2008; Morgan and Rubin, 2012).

We estimated the following model:

$$y_{ij} = \beta_0 + \beta_1 Treat_j + \beta_2 PreTest_{ij} + \beta_3 X_j + u_j + \varepsilon_{ij}$$

$$u_j \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where:

- individual i is nested in school j;
- y_{ij} is the score on the chemistry subscale of AQA GCSE Double Award Science;
- $PreTest_{ij}$ is the simple aggregation of English and mathematics KS2 scores;
- $Treat_j$ is our school-level treatment indicator; and

- X_j is a categorical variable indicating block assignment—a vector of stratification variables, as previously defined in this section.

As the model constant term, β_0 recovers the mean of the outcome variable when other covariates are set to zero/baseline categories, β_1 recovers our estimated treatment effect, u_j and ε_{ij} represent a school-level random effect and an error term at the pupil level respectively.

All models were estimated using Bayesian inference (Gelman et al., 2014) with the software rstanarm through R using weakly informative priors. These priors were set in the analysis code, available in Appendix I, rather than using default settings. Visual inspection of the MCMC traces, the R-hat diagnostic (Gelman and Rubin, 1992), and posterior predictive checks were used as model fit diagnostics. Uncertainty around the primary impact estimate is reported using 95% Bayesian credible intervals. We also report frequentist 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 (2018), we will use a ROPE (Region of Practical Equivalence) analysis set at an effect size of ±0.1 around 0 to examine whether the null hypothesis should be accepted as practically distinguishable from no effect. This procedure examines the proportion of the Highest Density Interval (HDI) that falls within the ROPE predetermined effect size. In essence, the ROPE indicates the proportion of the parameter (effect size) values that are equivalent to the null value for practical purposes. This approach was selected because it provides comparability with a standard frequentist approach and effect sizes (Kruschke, 2018) used in reporting other EEF trials.

We also report the results of a frequentist approach, for consistency with other EEF trials, as set out in the section entitled Additional Analyses below.

See Appendix I for syntax used for analysis.

Secondary analysis

The secondary analyses focused on the four secondary outcomes: the GCSE science total score and the three AOs for chemistry. We modelled each of these separately using the same model specification to the primary outcome, except for substituting the outcome variable as appropriate.

Analysis in the presence of non-compliance

The effect of school-level compliance was analysed using an Instrumental Variables (IV) approach with group allocation as the instrumental variable for the compliance indicator. We adapted our approach to compliance as a deeper understanding of the difficulties in delivering an ideal implementation of the SMART Spaces Revision intervention was developed. Capturing these difficulties was rather more complicated than originally envisaged, for example, due to higher variation in school's timetabling of science lessons than expected, hence we adopted a more nuanced approach to compliance. In this section, we set out first our agreed approach before noting our original compliance plan. We will report the results of both sets of analyses specified below.

As specified in the SAP, we agreed a set of compliance indicators with the developer, based on attendance at CPD and coaching sessions and the delivery of SMART Spaces lessons, to create a continuous compliance variable, $Comply_j$, specified as follows:

- Percentage of teachers attending CPD. This was calculated using the number of teachers who planned to deliver SMART Spaces Revision, dividing actual CPD and coaching sessions attended by total possible CPD and coaching sessions attended.
- Percentage of SMART Spaces lessons taught. We calculated this using the actual number of lessons delivered divided by the total lessons expected to be delivered, that is, six SMART Spaces lessons multiplied by the number of double award classes.
- Percentage of delivery with appropriate spacing. This was calculated as the number of classes receiving the SMART Spaces Revision intervention with appropriate spacing divided by the total number of double award classes.

Attendance registers from training and coaching sessions were collected from the developer in order to assess attendance. The teacher survey was used to collect information about the number of SMART Spaces lessons taught and the implementation of spacing. We estimated a (first stage) model of compliance, with $Comply_j$ as the dependent variable, treatment assignment as the instrumental variable, and block assignment and pre-test score as additional independent variables.

$$Comply_j = \beta_0 + \beta_1 Treat_j + \beta_2 PreTest_{ij} + \beta_3 X_j + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

The predicted values of $Comply_j$ from the first stage are used in the estimation of the second stage model of our outcome measure, from which the instrumental variable is excluded.

$$y_{ij} = \beta_0 + \beta_1 \widehat{Comply}_j + \beta_2 PreTest_{ij} + \beta_3 X_j + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

Reliable packages for implementing a multilevel or Bayesian approach to instrumental variables were unavailable within the SRS. Hence, in a change to the approach prespecified in the SAP, the compliance analysis was implemented in a frequentist two-stage least squares model using lfe (Gaure et al., 2022).

Our primary outcome of interest from the second stage model is the coefficient for $\widehat{Comply}_j$, where $\widehat{Comply}_j$ are the predicted values of treatment receipt derived from the first stage model. This coefficient recovers the estimated effect of the intervention among compliers (that is, those who would not have implemented SMART Spaces Revision in the absence of being assigned to treatment, but did so because of that assignment). Results for the first stage will be reported alongside (i) the correlation between the instrument and the endogenous variable and (ii) an F test as per EEF (2018) statistical analysis guidance.

In our original compliance plan, we agreed a definition of minimum compliance with the developer based on attendance at training and coaching sessions and the delivery of SMART Spaces lessons to create $IdealComply_j$, a binary compliance variable specified as:

- all (100%) of teachers delivering SMART must receive training (assessed through QUB/HTSA records of attendance); and
- all double award classes to receive both sets of three SMART Spaces lessons (six sessions in total) delivered with appropriate ‘spacing’ as set out in the protocol (assessed through the teacher and pupil survey).

As the developer and evaluation teams appreciated the complexity of delivering an ideal implementation, we had concerns that this plan may have led to underestimating the local average treatment effect for minimum compliance. For this reason, just prior to publication of the SAP version 1.0, we adjusted our plan as set out above. Nevertheless, for the purposes of transparency, we also carried out our original compliance plan, as an additional robustness check, using an identical approach as for the continuous compliance variable.

We also investigated the effects of ‘non-compliance’ in the control group. The SMART Spaces Revision intervention is not publicly available, so schools in the control group did 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 report these ‘always compliers’ using survey data. The survey data was not judged to be sufficiently robust to investigate this quantitatively.

Missing data analysis

We conducted a missing data analysis on all models because the threshold (as set out in the SAP) of more than 5% of data was missing. We first assessed whether it is plausible to assume that the missing data is missing at random (MAR), since this is a prerequisite for missing data modelling to produce meaningful results. To do this we created an indicator variable for each variable in the impact model specifying whether the data is missing or not. We then used logistic regression to test whether this missing status can be predicted from the following variables: all variables in the primary outcome analysis model plus eligibility for FSM (and proportion eligible for FSM in the school), sex of the pupil, and GCSE science raw score. Predictability was confirmed, so we proceeded to the appropriate next step of this strategy.

For situations for which the MAR assumption appears to hold and only the outcome variable in the model was missing, we re-estimated the treatment effect using our prespecified model with the addition of the covariates found to be statistically significantly predictive of missingness of the outcome.

For situations for which the MAR assumption appears to hold and any variable other than the outcome variable in the model is missing, we used all variables in the analysis model plus eligibility for FSM (and proportion eligible for FSM in the school), sex of the pupil, and GCSE science raw score to estimate a Multiple Imputation (MI) model using a fully conditional specification, implemented using multiple imputation by chained equations (mice) to create 20 imputed datasets. Due to time constraints in accessing and running a Bayesian analysis in the ONS SRS, in a change from the Bayesian approach specified in the SAP, this was conducted using a frequentist approach and implemented using the mice imputation package within R (van Buuren and Groothuis-Oudshoorn, 2011). We re-estimated the treatment effect using each dataset and calculated the average treatment coefficient, standard error, and p-value using Rubin's (2004) combination rules.

Subgroup analyses

A subgroup analysis was carried out for everFSM pupils, adding an interaction effect between treatment and everFSM to the primary outcome model. For this everFSM subgroup analysis, we estimated the following model:

$$y_{ij} = \beta_0 + \beta_1 Treat_j + \beta_2 PreTest_{ij} + \beta_3 X_j + \beta_4 Treat_j * FSMever_{ij} + \beta_5 FSMever_{ij} + u_j + \varepsilon_{ij}$$

$$u_j \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where the variables are the same as in the primary outcome analysis with the addition of $\boldsymbol{FSMever_{ij}}$, an indicator of FSM eligibility available from the National Pupil Database (NPD), and $\boldsymbol{Treat_j * FSMever_{ij}}$ is an interaction between the school-level treatment indicator and everFSM. This analysis required the primary outcome data to be matched with the NPD to provide 'everFSM' data for pupils. NPD data was matched to original pupil data collected before randomisation.

Secondly, and similarly, a subgroup analysis was carried out for sex, adding an interaction effect between treatment and sex to the primary outcome model. This subgroup analysis is deemed necessary 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). For this sex subgroup analysis, we estimated the following model:

$$y_{ij} = \beta_0 + \beta_1 Treat_j + \beta_2 PreTest_{ij} + \beta_3 X_j + \beta_4 Treat_j * Sex_{ij} + \beta_5 Sex_{ij} + u_j + \varepsilon_{ij}$$

$$u_j \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where the variables are the same as in the primary outcome analysis with the addition of $\boldsymbol{Sex_{ij}}$, a binary indicator of the sex of the pupil, and $\boldsymbol{Treat_j * Sex_{ij}}$ is an interaction between the school-level treatment indicator and sex.

Finally, a subgroup analysis was carried out to investigate differential treatment effects depending on prior attainment at pupil level by adding an interaction effect between treatment and $\boldsymbol{PreTest_{ij}}$ to the primary outcome model. As noted in the evaluation protocol, one of the findings of the optimisation study was that teachers considered the intervention to have greater benefits for lower attaining pupils and that higher attaining pupils were less engaged and perceived there to be less benefit (O'Hare, 2017, p.32). For this pupil-level, prior-attainment subgroup analysis, we will estimate the following model:

$$y_{ij} = \beta_0 + \beta_1 Treat_j + \beta_2 PreTest_{ij} + \beta_3 X_j + \beta_4 Treat_j * PreTest_{ij} + u_j + \varepsilon_{ij}$$

$$u_j \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where the variables are the same as in the primary outcome analysis with the addition of $\boldsymbol{Treat_j * PreTest_{ij}}$, an interaction between the school-level treatment indicator and pupil-level prior attainment. Prior attainment subgroups were defined by tertiles based on prior attainment of the overall sample.

For each subgroup analysis, if a significant interaction had been found, we would have run a separate model using only the relevant subgroup as the estimation sample, using the same model as our primary analysis. However, this eventuality did not occur.

The subgroup analyses were conducted for both the primary and secondary outcomes, with minor model adaptations as necessary.

Additional analyses and robustness checks

The variation in schools size might affect the results because, for example, larger schools might have more specialist chemistry teachers. Hence, we conducted a sensitivity analysis with cluster size $\boldsymbol{ClusterSize_j}$ included as an additional covariate in the primary and secondary outcome models to take account of variation in cluster size. That is, we estimated the following model:

$$y_{ij} = \beta_0 + \beta_1 Treat_j + \beta_2 PreTest_{ij} + \beta_3 X_j + \beta_4 ClusterSize_j + u_j + \varepsilon_{ij}$$

$$u_j \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where $\boldsymbol{y_{ij}}$ is variously the primary and secondary outcome measures.

O'Hare et al. (2017) found that pupil engagement in the intervention was a significant implementation factor, with higher engagement score predicting more positive outcome change. As part of the IPE, we explored the effect of engagement on the intervention using a pupil-level engagement measure based on survey items from the optimisation study (OS).

We have used factor analysis as an exploratory tool for validating this measure using data from the OS. We have also employed Rasch analysis as a means of confirming the factor analysis results and to construct an interval measure of engagement in spaced learning from these items.[7] We conducted an interaction analysis, by estimating the following model:

$$y_{ij} = \beta_0 + \beta_1 Treat_j + \beta_2 PreTest_{ij} + \beta_3 X_j + \beta_4 Treat_j * Engage_{ij} + \beta_5 Engage_{ij} + u_j + \varepsilon_{ij}$$

$$u_j \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where the variables are the same as in the primary outcome analysis with the addition of $\boldsymbol{Engage_{ij}}$ (a measure of pupil-level engagement) and $\boldsymbol{Treat_j * Engage_{ij}}$ (an interaction between the school-level treatment indicator and pupil-level engagement measure). In the first instance, we used the raw score on engagement items as the measure of pupil-level engagement before subsequently performing this analysis with the Rasch-constructed scale.

As a robustness check, we fitted each model in a frequentist framework using the lme4 package in R. This also provides consistency with other EEF trials. Our original intention was to additionally replicate this classical analysis using the alternative software MLwiN. Due to difficulties accessing this software in the ONS SRS, this additional analysis was not carried out.

---

[7] For details of the validation process, see the SAP.

Estimation of effect sizes

Effect sizes were calculated using the Cohen's *d* ES for cluster randomised trials as per the current EEF (2018) statistical analysis guidance for evaluations. The formula is specified below:

$$ES = \frac{\bar{Y}_t - \bar{Y}_c}{\sqrt{(\sigma_u^2 + \sigma_\varepsilon^2)}}$$

where $\bar{\boldsymbol{Y}}_{\boldsymbol{t}} - \bar{\boldsymbol{Y}}_{\boldsymbol{c}}$ is recovered from $\boldsymbol{\beta_1}$ in the primary ITT analysis (and, thus, adjusted for pre-test and stratification variables). $\boldsymbol{\sigma_u^2}$ represents the variance of the school-level random effects and $\boldsymbol{\sigma_\varepsilon^2}$ the variance of the pupil-level random effects in the primary ITT analysis.

More specifically, we use Cohen's *d* ES for two reasons as follows:

For multilevel models, such as those used in this evaluation, the EEF (2018, p.4) guidance recommends using Cohen's *d*. Furthermore, there is negligible difference between Hedges' *g* and Cohen's *d* in a trial of this size: as per the EEF (2018, p. 4, footnote 10) guidance stating that 'the difference between Hedges' g and Cohen's d is minimal for samples over 30 so either could be used in practice'.

As per Adkins (2017), we computed effect sizes directly in Stan. In lme4, the sim() function from the Applied Regression Modelling package (arm) in R was used to compute the classically derived estimates using the same methodology (Adkins, 2017).

Estimation of ICC

We employed a random, intercept-only, multilevel model to estimate the intra-cluster correlation (ICC) of the pre- and post-tests at school level, as follows:

$$y_{ij} = \beta_0 + u_j + \varepsilon_{ij}$$

$$u_j \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where individual $\boldsymbol{i}$ is nested in school $\boldsymbol{j}$, $\boldsymbol{y_{ij}}$ is the score on the chemistry subscale of AQA GCSE Double Award science, $\boldsymbol{\beta_0}$ represents the grand mean for the outcome variable, $\boldsymbol{u_j}$, and $\boldsymbol{\varepsilon_{ij}}$ represent a school-level random effect and an error term at the pupil level, respectively.

The ICC itself was estimated from this model using the following equation:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$$

Implementation and process evaluation

The IPE adopted a mixed methods approach involving survey and case studies as described below and was conducted in accordance with the EEF's IPE guidelines.

Data collection

Data collection involved 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 its implementation. Hence, in

and

, we set out how this data is linked to the logic model.

Surveys

Surveys were administered on paper using optical character recognition (OCR) or online technology. See Appendix J for full copies of all surveys.

We collected the following data for the intervention group.

- SMART lead teacher survey

  This survey was designed to be completed by either the head of science, head of chemistry, or the designated SMART lead teacher in each intervention school. It was administered online after the GCSE examination to gather data on their perceptions of the intervention and its implementation, how implementation was supported, and the quality and variability of delivery. In some intervention schools, the head of science or chemistry delegated the leadership role for SMART to another member of staff, the designated SMART lead teacher. In these cases, the survey was normally completed by that teacher. The survey was piloted 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.][8]

- Teacher survey

  This survey was designed to be completed by all teachers in intervention schools. It was administered online after the GCSE examination 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 pupil engagement. The survey was piloted on a sample of teachers (N = 25) before use after feedback from the developers and from teachers in the parallel SMART Space Teaching pilot. [Fidelity, (teacher) responsiveness, adaptation, reach, quality, differentiation, implementation factors, RQ6, RQ7, RQ8, RQ9, RQ10, RQ11, RQ12, RQ13, RQ17.]

- Pupil survey

  This survey was designed to be completed by all pupils in intervention schools and classes. It was adapted from the implementation survey used in the pilot to gather data on pupil engagement, their general revision practices for school science, their approaches to revision at home, and their self-concept of, and beliefs about, chemistry. It was administered using machine readable paper surveys at the end of the pupils' final SMART Spaces lesson. We adapted and then validated this survey in three rounds of piloting using factor analysis and Rasch modelling to validate the engagement scale (see Appendix J for further details). Some additional items were added to gauge school revision practices, informed by previous work on pupil ratings of instruction (for example, Nitz et al., 2014). [Fidelity, dosage, (pupils) responsiveness, reach, adaptation, RQ6, RQ14, RQ15.]

We collected the following data for the control group.

- Control group teacher survey

  This survey was designed to be completed by teachers within each control school. These investigated schools' revision practices and whether schools use spacing (or interleaving) approaches (and thus allow us to understand better the counterfactual) and included items common to the intervention group teacher survey for purposes of comparability. It was administered online after the GCSE examination (as per the teacher surveys for the intervention group). Our original plan was to have a separate survey to be completed by the head of science or chemistry in each control school but it was decided to combine the two surveys into one because of high overlap across items and to reduce burden on control schools. [Monitoring control group, RQ16.]

- Control group pupil survey

  This survey of a sample of pupils in the control schools (nSCon~150) aimed to establish current revision practices in schools and their approaches to revision at home and included items common to the intervention group pupil

[8] Terms in square brackets cross-reference to research questions (RQs) and to criteria from Humphrey et al.'s (2016), Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook.

survey for purposes of comparability. The samples of pupils were selected by the SMART lead teacher in each school. It was administered using machine readable paper surveys just prior to the GCSE examination (as per the pupil surveys in the intervention group). We validated this survey using statistical techniques (for example, Rasch modelling) and further piloting. This validation is reported in the SAP. [Monitoring control group, implementation, RQ16.]

Sample size and response rates for the surveys are shown in Table 4.

Table 4: IPE surveys—sample sizes

Survey	Achieved Sample	Total	Response Rate
SMART lead teacher survey	53	54	98%
SMART teacher survey	99	224	44%
SMART teachers and lead teachers survey combined (schools)	152 (53)	278 (54)	55% (98%)
SMART pupil survey	3694*	6465	57%
Control group teacher survey (schools)	293 (71)	350 (70)	84% (99%)
Control group pupil survey	5793	7633	76%

* The engagement scale was created on the basis of all pupils in the sample who had responded to all the engagement items (N = 2437).

Case Studies

Case studies were used to supplement the survey data and to elucidate the processes and evaluation. [Fidelity, responsiveness, adaptation, reach, quality, implementation factors, RQ6, RQ7, RQ8, RQ9, RQ11, RQ12, RQ13, RQ14, RQ17.] We conducted six case studies of school implementation. Schools were selected purposefully in order to cover a range of levels of school engagement with the SMART Spaces Revision programme in addition to school type, average KS2 score, and FSM proportion. Details of the case study schools are shown in Table 5.[9]

Table 5: Characteristics of case study schools

	A	B	C	D	E	F
OFSTED	N/A	Good (2016 and 2022)	Good (2018)	Outstanding (2014)	Outstanding (2012)	Outstanding (2017)
Urban / Rural	Rural	Rural	Urban	Urban	Urban	Urban
School type	Academy sponsor-led	Foundation	Foundation	Academy converter	Academy converter	Academy converter
Average KS2 score (based on randomisation data)	4.18	4.16	3.90	4.52	4.30	4.11
FSM proportion*	25%	20%	15%	10%	10%	20%
School overall size (roll)	1400	1000	1200	1300	1700	1000
GCSE Combined Science Cohort (at randomisation)	190	100	130	140	90	120
Level of engagement with SMART**	Medium	High	High	Low	High	Medium

[9] For the choice of case study schools, school engagement was assessed on the basis of discussion with the developers.

* FSM proportion: for pseudonymity and data protection reasons, FSM proportion, school size, and Combined Science cohort size are rounded to nearest 5%, 100, and 10, respectively.
** The level of engagement is based on the compliance measure, which was on a scale from 0 to 1: high > 0.67; medium 0.50 to 0.66; low < 0.50.

Case study data was collected via:

- observation of SMART training sessions—to investigate fidelity to training, as set out in the TIDieR framework, and the role of training in supporting teachers;
- observation of teacher feedback in practice (coaching) sessions and a brief interview with trainers—to investigate fidelity to coaching, as set out in the TIDieR framework, and the role of coaching in supporting teachers;
- observation of two of the three intervention spaced lessons—two times A-B-C or two times D-E-F; we observed the whole lesson where the timetable allowed and, in other cases, observed parts of lessons; at least one entire lesson, and parts of at least two lessons, were observed for each teacher—to investigate factors influencing fidelity to the teaching model, as set out in the TIDieR framework, as well as responsiveness, adaptation, reach, quality of delivery, and pupil engagement;
- brief interviews with each case study teacher—to investigate implementation factors and teachers' views about the approach more generally;
- interviews with head of science or SMART lead teacher—to investigate implementation factors and teachers' and schools' views about the approach more generally; and
- focus group interviews with pupils (these were additional to the planned IPE as set out in the evaluation protocol)—to investigate pupil engagement.

Observations followed a predetermined protocol and interviews were semi-structured. See Appendix J for interview schedules; the case study data collected is summarised in Table 6.

Table 6: Summary of data collected from case study schools

	A	B	C	D	E	F	Total
Observation SMART training session	Yes	Yes	No	No	No	Yes	3
Observation of teacher coaching sessions	Yes: 3 teachers	Yes: 4 teachers	Yes: 2 teachers	Yes: 2 teachers	No	Yes: 3 teachers	14
Observation of SMART lessons	3 full lessons x 2 days	3 full lessons x 2 days	5 parts of double lessons x 2 days	3 full lessons x 2 days	3 full lessons x 2 days	3 full lessons x 1 day	37
Interviews with SMART lead teachers	Yes	Yes	Yes	Yes	Yes	Yes	6
Interviews with other teachers	2 teachers (individual)	2 teachers (group)	2 teachers (individual)	3 teachers (group)	3 teachers (group)	1 teacher	13
Focus group interviews with pupils	Yes	No	Yes	Yes	Yes	No	4

Note that four other schools were initially recruited as case study schools. Schools G, H, and J left the trial after our initial case study visits to observe training, although they submitted GCSE data. School K left before the initial visit but also submitted GCSE data. In responding to this we recruited further case study schools, which is why we missed observation of training for schools C, D, and E. We over-recruited case study schools to compensate, which is why we conducted six case studies rather than the planned five.

Interviews with the developers (QUB, HTSA) were conducted in order to better understand the intervention and thus inform the IPE data collection. [Fidelity, adaptation, implementation factors, RQ6, RQ8, RQ9, RQ16, RQ17.]

We planned to conduct five short telephone interviews with heads of science or chemistry in control schools but, due to time and scheduling constraints, these were not conducted.

Implementation and process evaluation data analysis

Questionnaire and survey data

Surveys were analysed descriptively and, where direct comparisons could be made between the intervention and control groups, using inferential statistics (using independent samples t-tests). For the purposes of this IPE, we judged differences in Cohen's *d* of > 0.1 to be potentially significant and, thus, please note that where we refer to statistical significance these results should be seen as indicative rather than conclusive (McDonald, 2014); we also did not consider the clustering of pupils in schools, hence, due to the large pupil sample size, some small effects were statistically significant on our tests but not judged to be of practical significance.

Case study and interview data

The case study data and interviews were analysed thematically using Braun and Clarke's (2006) six-stage process and informed by the survey results.

Analysis of materials

In order to further assess quality, we analysed the revision material relative to the AQA chemistry specification, with a particular focus on coverage of curriculum. This was conducted using expert judgment. [Quality, programme differentiation, RQ12, RQ16.]

Assessment of usual practice

We assessed 'usual' practice at baseline via the head of science and teacher surveys for intervention and control schools and in case studies for intervention schools. Given the focused and well-defined nature of this intervention (revision in the run-up to a GCSE), we found that teachers had reliable recall of revision practices in previous years and, hence, that the endpoint surveys were sufficient to capture baseline practice.

Table 7 provides a summary of the how the data and RQs relate to the IPE dimensions and implementation factors.

Table 7: IPE methods overview

Research methods	Data collection methods	Participants/data sources	Data analysis methods	Research questions addressed	Implementation/ logic model relevance
SMART lead teacher survey	Online survey	SMART Lead Teachers / Head of Science/chemistry	Descriptive statistics	6, 7, 8, 9, 11, 12, 13, 17	Fidelity, (schools) responsiveness, adaptation, reach, differentiation, implementation factors
Teacher survey	Online survey	Intervention Teachers (including SMART Lead Teachers)	Descriptive / inferential statistics	6, 7, 8, 9, 10, 11, 12, 13, 17	Fidelity, (teacher) responsiveness, adaptation, reach, quality, differentiation, implementation factors
Pupil survey	Paper survey	Pupils in intervention classes	Descriptive / inferential statistics	6, 14, 15	Fidelity, dosage, (pupils) responsiveness, reach, adaptation
Control group teacher survey	Online survey	Teachers & Heads of Science in control group schools	Descriptive / inferential statistics	16	Monitoring control group
Control group pupil survey	Paper survey	Pupils in control group	Descriptive / inferential statistics	16	Monitoring control group, implementation

Case studies	Observations of lessons, training sessions, Interviews	Teachers, SMART Lead Teachers, Pupils from intervention schools	Thematic analysis	6, 7, 8, 9. 11, 12, 13, 14, 17	Fidelity, responsiveness, adaptation, reach, quality, implementation factors
Interview with SMART developers	Focus group interviews	QUB and HTSA teams, separately	Thematic analysis	6, 8, 9, 16, 17	Fidelity, adaptation, implementation factors
Analysis of SMART materials	Materials supplied by developer	SMART Powerpoints and handbooks	Expert judgment	12, 16	Quality, programme differentiation

and

indicate how data collected links to key elements within the logic model.

Figure 7: SMART Spaces Revision logic model—overall, with IPE measures

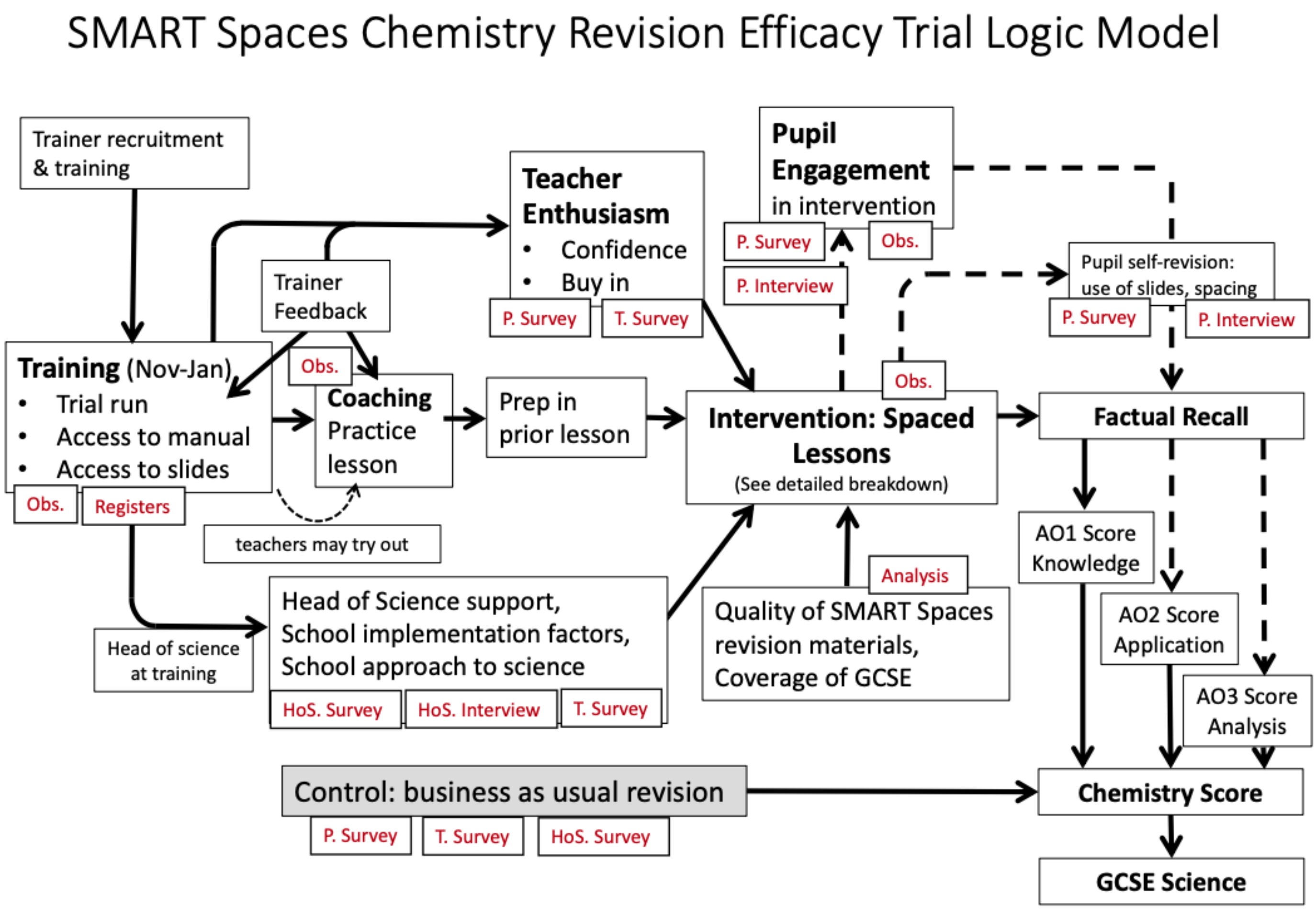
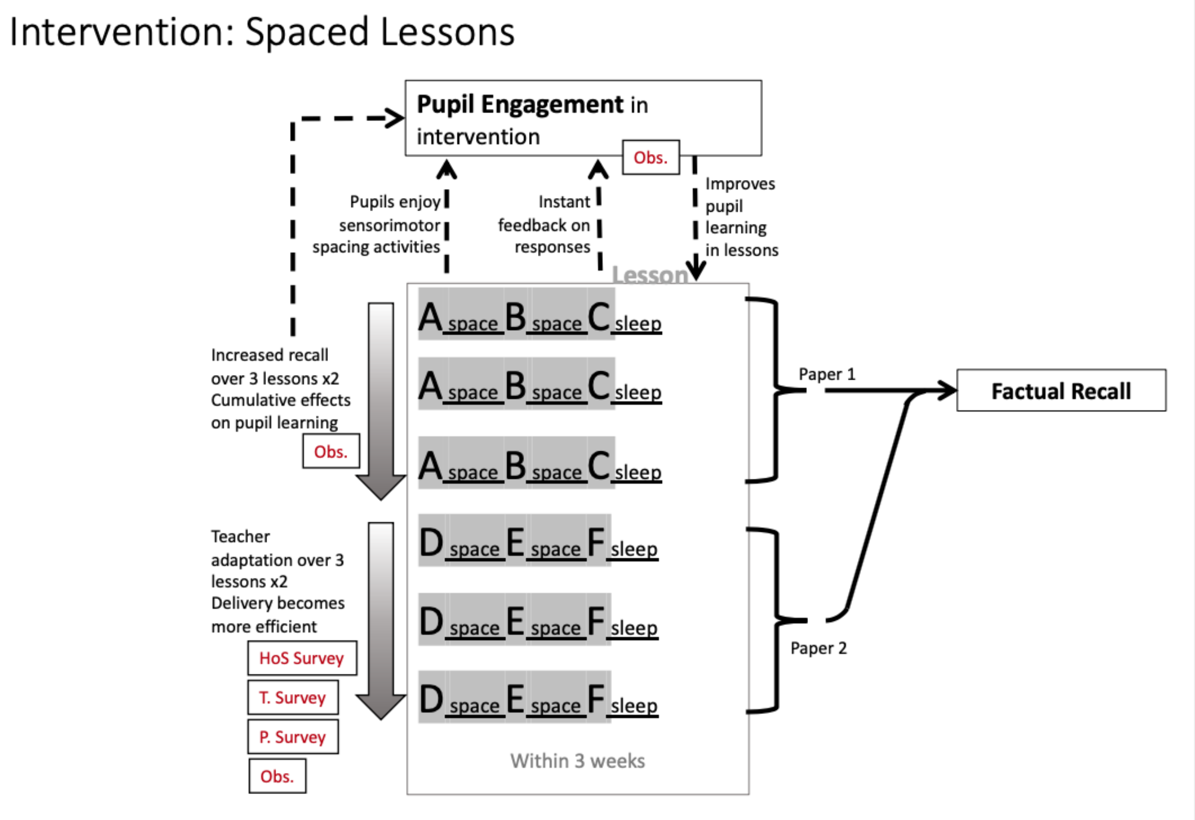


Figure 8: SMART Spaces Revision logic model—intervention spaced lessons element, with IPE measures



Costs

Information about the costs associated with the intervention were gathered from the developer. All resources required for the intervention (for example, manuals and spacing activity resources) were supplied by the developer. Initial training and follow-up visits involved twilight sessions and, hence, required no supply cover. The teachers from some schools attended another local school for the training day. We have assumed the travel costs to be minimal. Additional qualitative data on time and cost was collected through interviews as part of case studies. Determining the overall cost per pupil per year was in accordance with the EEF (2019) guidance on cost analysis.

Timeline

Table 8 shows a timeline of activities.

Table 8: Timeline of activities

Dates	Activity	Responsible/ leading
Mar–Dec 2018	Recruitment	QUB/HTSA
May–Jun and Sep–Oct 2018	Initial data collection, pre-randomisation	UCL
Oct and Dec 2018	Randomisation (two stages)	UCL
Dec 2018–Mar 2019	SMART Spaces training (and observations for case studies by UCL)	HTSA/QUB
Jan–Apr 2019	SMART Spaces coaching (and observations for case studies by UCL)	HTSA/QUB
23 Apr–10 May 2019	SMART Spaces Revision intervention delivery (and case study data collection by UCL)	HTSA/QUB

23 Apr–10 May 2019	Pupil survey administered during final SMART Spaces lesson as last spacing activity	UCL
16 May and 12 Jun 2019	GCSE science examinations	-
Jul–Aug 2019	Online teacher survey (intervention and control groups)	UCL
Aug–Sep 2019	Data collection (UMS scores)	UCL

Impact evaluation results

Participant flow including losses and exclusions

The flow of participants is detailed in Figure 9. Of the approximately 3,250 schools that were approached, 299 submitted expressions of interest. However, only 231 of these schools returned signed MOUs; of those that did so, 88 were excluded including 60 schools that were ineligible to take part because they did not use the AQA examination board for the Combined Science GCSE award or because the school was not in England. A further 18 schools subsequently withdrew from the project for reasons including the school moved to the parallel SMART Spaces Teaching Pilot (four schools), or due to concerns about data protection (one school), or staff capacity (nine). The remaining 125 schools were randomly allocated to the intervention and control groups, as described above.

Figure 9 shows that five schools were lost at follow-up stage—four intervention and one control. Three of these withdrawals were due to data protection concerns, one due to a school closing, and one was a treatment school that withdrew early in the trial without giving any reason.

The results of MDES calculations at various points of the trial are presented in Table 9. The MDES was 0.198 at the design stage and remained at this level at randomisation. The MDES decreased to 0.163 at the analysis stage, thus increasing the power of the trial. This was due to a much higher than anticipated school-level correlation between KS2 and GCSE chemistry scores.

Figure 9: Participant flow diagram for the primary analysis

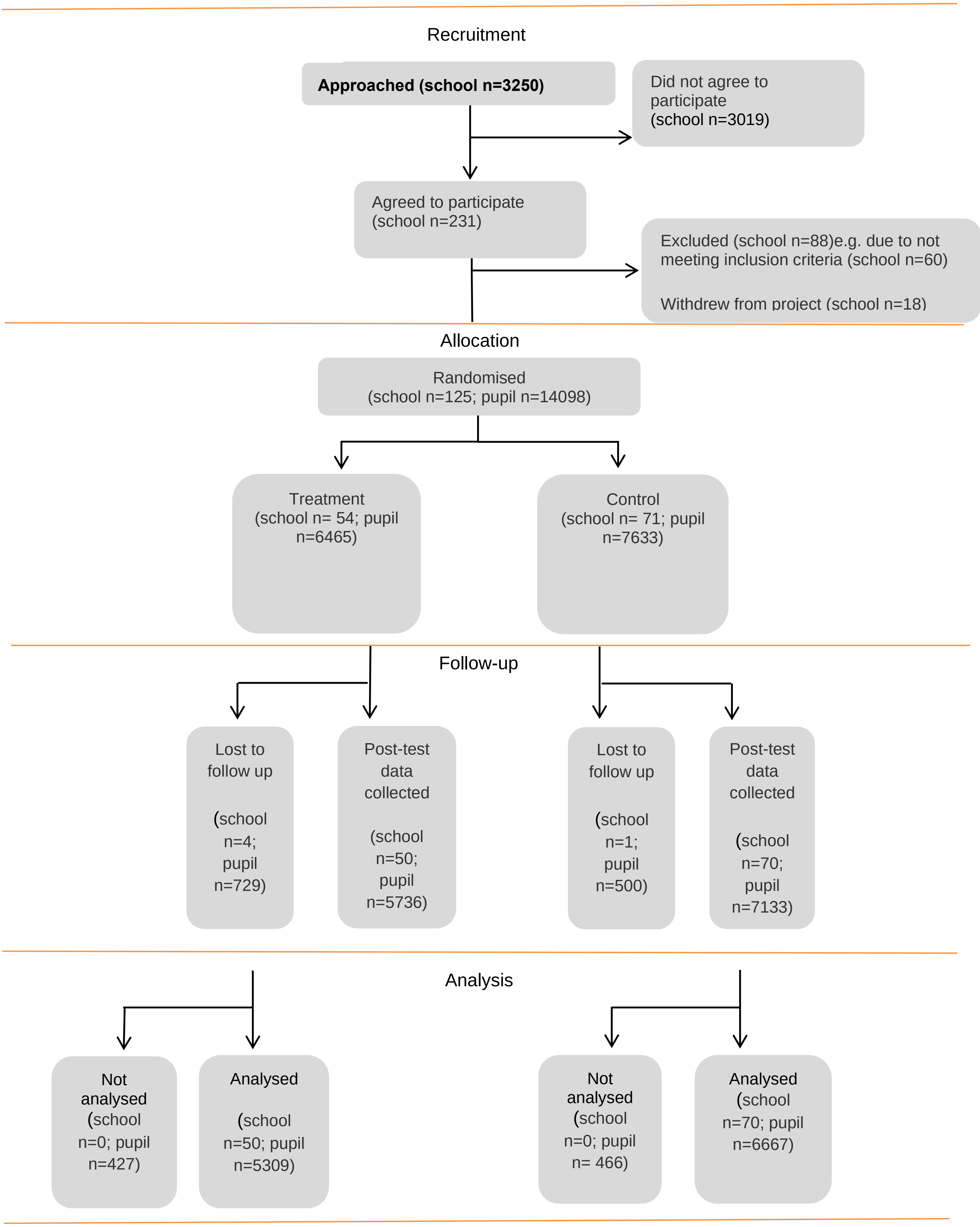


Table 9: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	**FSM**	**Overall**	**FSM**	**Overall**	**FSM**
MDES		0.198	0.210	0.196	0.209	0.163	0.169
Pre-test/post-test correlations	Level 1 (pupil)	0.50	0.50	0.50	0.50	0.586	0.549
	Level 2 (school)	0.25	0.25	0.25	0.25	0.702	0.552
Intracluster correlations (ICCs)	Level 2 (school)	0.15	0.15	0.15	0.15	0.178	0.113
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		100	25	113	25	99.8	26.4
Number of schools	Intervention	50	50	54	54	50	50
	Control	75	75	71	71	70	69[10]
	Total:	125	125	125	125	120	120
Number of pupils	Intervention	5000	1250	6465	1796	5309	1466
	Control	7500	1875	7633	1837	6667	1676
	Total:	12500	3125	14098	3633	11976	3142

Attrition

Attrition rates are shown in Table 10. Overall attrition at pupil level from randomisation to analysis was 15.1%. Approximately 9% of this was due to school withdrawals as described under Participant Flow above. The remaining attrition was largely due to trial pupils having missing KS2 data in the NPD. In addition, small numbers of pupils could not be matched to the NPD or withdrew their data from the research. There was some evidence of differential attrition between the intervention and control groups (5.2%pts). Only 1.6%pts of this imbalance was due to the different levels of school withdrawal across the two groups.

[10] One school had no FSM pupils in the cohort of GCSE Combined Science pupils who did not withdraw their data from the study.

Table 10: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	6465	7633	14098
	Analysed	5309	6667	11976
Pupil attrition (from randomisation to analysis)	Number	1156	966	2122
	Percentage	17.9	12.7	15.1

Pupil and school characteristics

Table 11 presents the baseline characteristics of intervention and control schools and pupils as randomised.

The balance at pupil level for KS2 attainment is good and within acceptable levels. There are some indications of imbalance for FSM, sex, and cluster size. At school level, in general, intervention and treatment schools are reasonably well-balanced in terms of Ofsted grading, setting (urban or rural), and school type (academy, community, and other). There are some indications of imbalances in terms of prior attainment, FSM, and the size of the school's Combined Science cohort with the intervention schools having slightly lower prior attainment, higher FSM proportions, and larger Combined Science cohorts compared to the control schools. This may be due to differences in schools' decisions about entry to the different GCSE Science awards. These imbalances are nevertheless relatively small and appear to be within acceptable levels. Overall, schools appear largely reasonably similar to the national averages in terms of setting (location), school type, and Ofsted grading, although the proportion of schools with a 'good' judgement is slightly lower than the national average (and consequently, the proportions of schools with 'outstanding' and 'requires improvement' or 'inadequate' judgements are slightly higher than the national average).

In order to assess whether these imbalances had any effect on the primary and secondary analyses, we conducted two additional robustness checks as reported in the Additional Analyses and Robustness checks sections below.

Table 11: Baseline characteristics of groups as randomised

		Intervention		Control		
School level (categorical)	**National-level mean**	**n/N (missing)**	**%**	**n/N (missing)**	**%**	
Ofsted: Outstanding*	14.5%	14 / 54 (2)	25.9	18 / 71 (0)	25.4	
Good	57.1%	25 / 54 (2)	46.3	37 / 71 (0)	52.1	
Requires imp./ inadequate	18.3%	13 / 54 (2)	24.1	16 / 71 (0)	22.5	
Setting: Urban	85.7%	44 / 54 (0)	81.5	56 / 71 (0)	78.9	
Rural	14.3%	10 / 54 (0)	18.5	15 / 71 (0)	21.1	
School type: Academy	72.8%	37 / 54 (0)	68.5	45 / 71 (0)	63.4	
Community	10.9%	9 / 54 (0)	16.7	9 / 71 (0)	12.7	
Other	16.2%	8 / 54 (0)	14.8	17 / 71 (0)	23.9	
School-level (continuous)		**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	
Average KS2 score	Not available	54 (0)	124.2 (10.1)	71 (0)	125.6 (11.4)	
FSM proportion	0.272	54 (0)	0.286 (0.153)	71 (0)	0.228 (0.139)	
Cluster size	Not available	54 (0)	119.7 (49.6)	71 (0)	107.5 (52.5)	
Pupil level (categorical)		**n/N (missing)**	**%**	**n/N (missing)**	**%**	
FSM		1905 / 6465 (72)	29.5	2016 / 7633 (87)	26.4	
Non FSM		4488 / 6465 (72)	69.4	5530 / 7633 (87)	72.4	
Sex, female		3286 / 6465 (43)	50.8	3585 / 7633 (60)	47.0	
Sex, male		3136 / 6465 (43)	48.5	3988 / 7633 (60)	52.2	
Pupil level (continuous)		**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**Effect size**
KS2 average score		6465 (576)	125.7 (32.0)	7633 (589)	126.4 (30.8)	0.024

* Nationally, 10.1% of schools are missing OFSTED data. These are mainly converter academies that have not had an inspection under their current status or free schools that are yet to be inspected.

Outcomes and analysis

Primary analysis

Table 12 presents the results of the analysis for the primary outcome measure (chemistry attainment at GCSE). It shows the unadjusted means for the intervention group (77.2) and the control group (75.7). Based on this, we calculate a Cohen's *d* effect size of -0.015 [95% Credible Interval: -0.17, 0.13], which is very close to zero (and is equivalent to zero months of progress). The region of practical equivalence was very large: 84% of the parameter (effect size) values ere equivalent to the null value for practical purposes. This provides little evidence of difference in the primary outcome (chemistry attainment) between the two groups. Hence, we found no evidence to support an effect of the SMART Spaces Revision intervention on the chemistry element of the GCSE combined award, as hypothesised in the logic model.

Table 12: Primary and secondary analysis

	Unadjusted means				**Effect size**		
	Intervention		**Control**				
Outcome	**n (missing)**	**Mean (95% Credible Interval)**	**n (missing)**	**Mean (95% Credible Interval)**	**Total n (intervention; control)**	**Cohen's *d* (95% Credible Interval)**	**ROPE %**
Primary outcome							
Chemistry subscale of GCSE Combined Science	5309 (1156)	77.17 (76.19, 78.10)	6667 (966)	75.74 (74.87, 76.56)	11976 (5309; 6667)	-0.015 (-0.171, 0.128)	84.4
Secondary outcomes							
GCSE Combined Science	5334 (1131)	217.01 (214.75, 219.27)	6706 (927)	215.49 (213.39, 217.47)	12040 (5334; 6706)	-0.033 (-0.185, 0.124)	81.0
AO1 score (chemistry knowledge)	5205 (1260)	32.04 (31.61, 32.44)	6442 (1191)	30.36 (29.99, 30.73)	11647 (5205; 6442)	0.061 (-0.087, 0.217)	69.6
AO2 score (chemistry application)	5205 (1260)	35.92 (35.46, 36.34)	6442 (1191)	35.95 (35.57, 36.34)	11647 (5205; 6442)	-0.051 (-0.190, 0.082)	77.2
AO3 score (chemistry analysis)	5205 (1260)	9.85 (9.69, 10.00)	6442 (1191)	10.01 (9.87, 10.14)	11647 (5205; 6442)	-0.067 (-0.195, 0.072)	69.5

Secondary analysis

The results of the secondary analyses are also presented in Table 12. The secondary outcomes were GCSE Science and the three assessment objectives (AOs) in chemistry.

As indicated in the logic model (Figure 3), AO1 chemistry knowledge is more closely aligned to the content, theory, and aims of the SMART Spaces Revision intervention in comparison to the primary outcome and the other secondary outcomes. Hence, one might expect to find a larger effect for this secondary outcome. In fact, this is the only outcome where we found a positive effect in favour of the intervention—0.06, equivalent to one month of progress. However, the region of practical equivalence was large (70%), which suggests that this is at best weak evidence of difference between the intervention and control groups. Thus, we found some weak evidence to support a small effect of the SMART Spaces Revision intervention on the AO1, chemistry knowledge, as hypothesised in the logic model.

The effect size for Combined Science GCSE was -0.03, which is equivalent to zero months of progress. The region of practical equivalence was very large (81%) providing strong evidence of no difference in this secondary outcome (science attainment at GCSE) between the two groups. Hence, we found no evidence to support an effect of the SMART Spaces Revision intervention on the GCSE Combined Science award, as hypothesised in the logic model.

The effect sizes for AO2 chemistry application and AO3 chemistry analysis were -0.05 and -0.07, respectively. These effect sizes are equivalent to one month of progress in favour of the control group. However, the regions of practical equivalence were large (77% and 70%, respectively) providing little evidence of practical difference in this secondary outcome (science attainment at GCSE) between the two groups. Hence, we found no evidence to support an effect of the SMART Spaces Revision intervention on the AO2 and AO3, as hypothesised in the logic model.

Analysis in the presence of non-compliance

As outlined in the Methods section, compliance was analysed at school level using an Instrumental Variables (IV) approach with group allocation as the instrumental variable for the compliance indicator. Two compliance variables were used: a continuous measure of compliance, the 'primary' compliance variable, and a binary indicator of 'ideal compliance'. In both cases, the analysis indicated no statistically significant effect for the compliance variable. We interpret these findings as not providing evidence of differential treatment effects among treatment schools with higher levels of compliance for the primary outcome measure, the GCSE chemistry subscale.

We investigated the uptake of spaced learning in control schools through the survey. This included a Likert scale item, 'How often do you use spaced learning in your chemistry teaching?', and also two parallel items for biology and physics. This suggested some, relatively infrequent, use of spaced learning in control schools. However, this is unlikely to be as systematic as the SMART Spaces Revision intervention and we do not judge it as a threat to the validity of the results.

See Appendix K for further details.

Missing data analysis

As outlined in the SAP, we proposed to implement our missing data strategy if more than 5% of data in the primary modelling is missing or if more than 10% of data for a single school is missing. Due to school withdrawals, both thresholds were met (see Table 10 and section on Attrition).

In order to assess whether it is plausible to consider this data as missing at random (MAR), we ran a logistic regression model to predict missingness in outcome data using all variables in the analysis model plus eligibility for FSM (and the proportion eligible for FSM in the school) and GCSE science raw score. The results of this analysis indicate that, pre-test, FSM eligibility and GCSE science raw score are statistically significant predictors of missing outcome data, which suggests this data is MAR. Sex was marginally predictive (p-value, < 0.1).

We first re-estimated the treatment effect for cases for which only the outcome variable was missing using the pre-specified model with the addition of FSM eligibility, GCSE science raw score, and sex (recalling pre-test is already included in the primary model). This analysis indicated no statistically significant effect for the intervention as in the main analysis: $d = 0.012$ (-0.11, 0.035).

Since the missing data threshold was also met for the predictor variables, we estimated a multiple imputation model using all variables in the analysis model plus FSM eligibility, sex, and GCSE science raw score. We created 20 imputed datasets and used these to re-estimate the treatment effect using Rubin's (2004) combination rules. This analysis indicated no statistically significant effect for the intervention as in the main analysis: $d = -0.037$ (-0.158, 0.084). See Appendix L for further details. Due to time constraints in accessing and running a Bayesian analysis in the ONS SRS, we did not conduct missing data analyses for the secondary outcomes as specified in the SAP.

Subgroup analyses

As is standard in all EEF-funded evaluations, we investigated whether there was evidence of differential effects among pupils eligible for FSM as a separate subgroup (n = 3142).[11] We started by considering an augmented version of the primary analysis model, including an interaction term between the treatment variable and membership of the FSM subgroup. The effect size estimate of this interaction (*d* = -0.07) was larger in magnitude compared to the main primary analysis. However, the region of practical significance was 83%, providing little evidence to suggest a differential effect. We repeated a similar analysis for each of the secondary outcomes and similarly found no evidence to indicate a differential effect. In addition, we conducted similar processes to investigate whether there was evidence of differential affects for sex and prior attainment. These analyses provided little evidence to suggest any differential effects. Table 13 presents a summary of the subgroup analyses.

Table 13: Subgroup analyses

	Outcome	**n**	**Cohen's *d*, ES for interaction**	**Cohen's *d*, 95% low CI**	**Cohen's *d*, 95% high CI**	**ROPE %**
FSM	Primary outcome: chemistry subscale	11976	-0.065	-0.142	0.018	82.5%
	GCSE Combined Award science	12040	-0.074	-0.154	0.004	75.3%
	AO1 score	11647	-0.086	-0.162	-0.001	64.9%
	AO2 score	11647	-0.078	-0.156	0.0002	70.4%
	AO3 score	11647	-0.024	-0.104	0.059	99.4%

	Outcome	**n**	**Cohen's *d*, ES for interaction**	**Cohen's *d*, 95% low CI**	**Cohen's *d*, 95% high CI**	**ROPE %**
Sex	Primary outcome: chemistry subscale	11976	-0.036	-0.102	0.037	99.5%
	GCSE Combined Award science	12040	-0.029	-0.100	0.034	100.0%
	AO1 score	11647	-0.033	-0.100	0.035	99.9%
	AO2 score	11647	-0.036	-0.102	0.038	99.5%
	AO3 score	11647	-0.039	-0.112	0.033	97.1%

	Outcome	**n**	**Cohen's *d*, ES for interaction**	**Cohen's *d*, 95% low CI**	**Cohen's *d*, 95% high CI**	**ROPE %**
Prior attainment	Primary outcome: chemistry subscale, tertile2	11976	-0.022	-0.105	0.059	99.3%
	Primary outcome: chemistry subscale, tertile3	11976	-0.025	-0.105	0.063	99.2%
	GCSE Combined Award science, tertile2	12040	-0.016	-0.097	0.064	100.0%
	GCSE Combined Award science, tertile3	12040	-0.026	-0.111	0.057	98.5%
	AO1 score, tertile2	11647	0.047	-0.037	0.125	91.7%

[11] Pupils eligible for FSM in the last six years (everFSM6) as is standard in EEF reports.

	AO1 score, tertile3	11647	0.064	-0.020	0.153	81.0%
	AO2 score, tertile2	11647	-0.034	-0.120	0.053	95.8%
	AO2 score, tertile3	11647	-0.018	-0.106	0.071	99.1%
	AO3 score, tertile2	11647	-0.073	-0.157	0.016	75.3%
	AO3 score, tertile3	11647	-0.091	-0.182	-0.001	58.4%

	Outcome	**n**	**Cohen's *d*, ES for interaction**	**Cohen's *d*, 95% low CI**	**Cohen's *d*, 95% high CI**	**ROPE %**
Cluster size	Primary outcome: chemistry subscale	11976	-0.009	-0.159	0.153	81.7%
	GCSE Combined Award science	12040	-0.027	-0.170	0.124	82.4%
	AO1 score	11647	0.064	-0.098	0.211	69.5%
	AO2 score	11647	-0.042	-0.177	0.104	80.2%
	AO3 score	11647	-0.062	-0.210	0.069	70.7%

Additional analyses and robustness checks

As outlined in the Methods section, we conducted two sensitivity analyses to assess the effect of imbalances in cluster size, school mean KS2 score, and school mean FSM proportion. We also ran analyses using frequentist methods. The results were not substantially different to the overall results; they provided no evidence to suggest any differential effects.

Finally, we conducted two interaction analyses to assess whether either pupil engagement in the intervention or examination entry were significant implementation factors. These indicated that neither higher pupil engagement (as measured by the pupil survey) nor examination entry were associated with higher (or lower) effects, suggesting that pupil engagement and examination entry were not implementation factors.[12]

See Appendix M for further details of these analyses.

Estimation of effect sizes

Effect sizes are calculated using Cohen's *d* and are presented in Table 12. None of the effect sizes estimated for this trial are of a substantive magnitude, nor are they statistically significant. This is the case for both the primary and secondary outcomes.

Estimation of ICC

The ICCs at school level for the primary and secondary outcomes are presented in

Table 14. These are similar to the estimate at the point of randomisation, 0.15.

[12] As would be expected, there was an effect for being in Higher compared to being in Foundation, that is, those in Higher did better than those in Foundation.

Table 14: ICC estimates

Outcome	Mean	SD	ICC
Is included in primary outcome			
Chemistry subscale	76.4	35.2	0.178
Secondary outcomes			
GCSE science total score	216.2	84.9	0.177
AO1 score	31.1	15.3	0.173
AO2 score	35.9	16.1	0.155
AO3 score	9.94	5.70	0.153

Implementation and process evaluation results

In this section, 'all teachers' (n = 152) refers to survey responses from both SMART leads (n = 53) and other teachers (n = 99). They were asked to report separately on each class that they taught: in total, 237 responses were received—SMART leads: 106; other teachers: 131. 'SMART leads' in this section refers to teachers who organised SMART Spaces Revision across the department. They were often, but not always, the head of department or head of chemistry.

Fidelity

For the purposes of this evaluation, we define 'fidelity' at teacher and class level as the extent to which teachers delivered the SMART Spaces Revision intervention and lessons as intended and described in the TIDieR framework and the logic model. This has some overlap with the compliance measure, which included the percentage of SMART Spaces lessons taught and of lessons delivered with appropriate spacing, although compliance was considered at a school level. For fidelity, we additionally focus on the content and pedagogic approach.

Fidelity was addressed in RQ6.

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

Broadly, teachers' fidelity to the intervention was generally good. When asked about their delivery of the SMART Spaces Revision to their classes (n = 237), teachers reported that they taught the full set of six SMART Spaces lessons in 66% of classes[13] with 60% of these lessons taking 60 minutes as intended. A further 10% and 18% taught at least five and four lessons, respectively, meaning 84% of classes received at least four (or two thirds) of the SMART Spaces lessons (and 76% received at least five of the six lessons). Teachers reported that in general they delivered the lessons as intended: they reported changing the language used in only 11% of classes, the content in only 3%, and the order in only 5%. However, teachers reported that they skipped some slides with 48% of classes, although this may be partly due to teachers skipping slides for pupils taking the foundation tier paper because some slides are aimed at pupils undertaking the higher tier examination paper. A higher proportion of teachers with less than five years' teaching experience (n = 67) reported that they changed the language used (19% vs 7%), Otherwise, there was little variation in this by subject specialism (chemistry or otherwise) of the teacher, or by experience. Evidence from the pupil survey also suggested that fidelity was reasonably good, although not as high as reported by the teachers: 51% of pupils (n = 3694) reported having six (or more) SMART lessons, 62% that their teacher covered all the slides in every lesson, and 77% that the idea behind spaced learning was well explained.

In terms of spacing within lessons, fidelity appeared to be high: 75% of all teachers (n = 152) reported that they taught all SMART Spaces lessons with the spacing as intended and 73% of pupils reported doing the required two spacing activities in each SMART lesson. Our case study visits also showed a high level of fidelity on the number of spaces: only one observed lesson (of the 34 observed) did not have the required two spaces. Fidelity to the spacing activities also appeared to be good. Juggling was suggested as the exemplar spacing activity, but the developer also provided guidance on alternatives. Teachers reported using juggling in 41% of classes. Of the other activities used, origami was the most commonly used (68%) followed by plasticine modelling (30%), balloon games (11%), and plate spinning (8%).[14] Teachers also reported that they used 'other' non-specified activities in 40% of classes, although this was lower for classes taught by SMART Leads than other teachers (26% vs 51%). Case study data suggests that these activities included colouring, throwing and catching, talking, and watching videos.

The intention was that each block of three SMART Spaces lessons should be taught on separate days over no more than a week. Evidence from the teacher survey suggests that teachers were able to achieve this with most classes:

[13] Aside from a small proportion of SMART lead teachers (3%), who may not have taught a GCSE combined award class, all the teachers taught at elast some SMART lessons to all classes, with 93% of all classes receiving three or more SMART lessons.
[14] Note that teachers were allowed to choose more than one activity as they could use different activities in different lessons so the percentages sum to greater than 100%.

teachers reported that, in 94% of classes, lessons were taught on separate days and for 66% of classes the block was taught over no more than a week.

Evidence from the pupil survey indicates that SMART Spaces lessons were delivered by a single teacher as recommended by the developers: 78% of pupils who responded to the survey (n = 3694) reported that their normal chemistry teacher delivered the SMART Spaces lessons.

These relatively high levels of fidelity suggest that fidelity at scale would be achievable. Indeed, 79% of SMART leads agreed or strongly agreed with the statement, ‘The SMART Spaces team provided everything we needed to run the intervention as a department.’ However, we note that the high quality of guidance and organisational support provided by the developer is likely to have been an additional factor in achieving fidelity within the trial. For example, by collecting information on when schools had timetabled SMART Spaces lessons, the developer’s interventions made it more likely that the SMART Spaces lessons would be timetabled and, hence, taught. Plans to scale up the intervention would need to give consideration as to how this support could be replicated.

Other aspects of implementation

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

The main potential barrier to implementation appears to be difficulties around the timetabling of the SMART Spaces lessons. While the evidence presented above under RQ6 (fidelity) suggests that most schools and teachers did manage to effectively timetable and deliver the lessons as required, it is nevertheless important to note that SMART lead teachers reported that they found that this was not straightforward, particularly for the second block of lessons for which 62% of leads disagreed or strongly disagreed with the statement, ‘It was easy to organise revision to fit the three SMART Spaces lessons covering AQA chemistry Paper 2 in.’

There are two aspects to timetabling: the length of the SMART Spaces lessons as designed and the need to fit three chemistry SMART Spaces lessons within a limited weekly number of science lessons. They are designed to be 60 minutes long but, in many schools, standard single lessons are less than 60 minutes. However, only a minority of teachers (around 20%) disagreed with the statement that the slides were ‘well-timed to fit within an hour’. Case study data suggests that schools with lessons shorter than 60 minutes managed to accommodate the SMART Spaces lessons in different ways. For example, one school simply moved more quickly through the content and the spacing activity while two other schools used part of double 50-minute science lessons to accommodate the 60-minute SMART Spaces lessons. In the cases where lessons were longer than 60 minutes, we also saw teachers adapt the intervention by slowing down the delivery of the revision materials, asking questions, and expanding on points so as to extend these periods of time beyond the designed 12 minutes.

Fitting three one-hour lessons devoted to chemistry within one week is a challenge and has potential knock-on effects for biology and physics, given a typical total weekly science allocation of four to five hours. However, this appeared to present a problem for only a minority of schools: 59% of SMART leads disagreed or strongly disagreed, and only 25% agreed or strongly agreed, with the statement: ‘There were tensions created between the chemistry revision using SMART Spaces Revision and revision of biology and physics.’ In three of the case study schools, lead teachers ‘borrowed’ revision lessons from biology, physics, and other subjects such as English and mathematics. This was aided by support from school leaders, a general flexibility around revision timetables, and the kudos of a research trial. These SMART leads all suggested that, in the long term, fitting the intervention around revision for other subjects might create challenges for implementation. In two further case study schools, the need to accommodate the SMART Spaces lessons impacted other school subjects, with lessons in other subjects (for example, maths and English) being given over to chemistry revision. This is possibly due to revision timetables across the school being different than normal teaching timetables throughout the year. Even in these cases it took considerable logistical work to accommodate the six SMART Spaces lessons and the support of senior leaders. In the cases where lessons were longer than 60 minutes, we also saw teachers adapt the intervention by slowing down the delivery of the revision materials, asking questions, and expanding on points so as to extend these periods of time beyond the designed 12 minutes.

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

Evidence from the case studies suggested that the designated SMART lead, who was not always the head of science or chemistry, played a key role in organising and managing the intervention as described above under RQ7. As well as facilitating the organisation of SMART Spaces lessons within the intervention, there is evidence from case studies that some lead teachers also played a key advocacy role in relation to the intervention. For example, in one case study school the head of department notes that the intervention is 'easy to sell' to colleagues because 'everything is prepared for them' (SMART Lead, School B). This view is reflected by other teachers in the school who see the intervention as providing high quality revision. In a further case study school, the head of department supported colleagues by dropping into intervention lessons and checking that everything was going well, and also discussing the intervention at team meetings.

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

The training provided by the developer was viewed positively by the majority of teachers: 70% of SMART leads and 82% of other teachers agreed or strongly agreed that the training provided them with 'everything I needed to deliver SMART Spaces'. The lower proportion of SMART lead teachers responding positively may be due to a lower proportion of these teachers attending the training (79%) in comparison to other teachers (94%). This may be due, in part, because a small number of heads of science responding to the survey did not teach SMART Spaces lessons.

Our observations of training sessions showed a very consistent format and delivery as delivered by a single member of the development team. We observed five training sessions with colleagues from seven schools in total, as some school colleagues travelled to nearby schools to attend training. The evidence behind the approach was explained and the processes of delivering the intervention both explained in detail and modelled. A majority of teachers responding to the survey (64%) agreed or strongly agreed that the evidence behind SMART Spaces Revision was important in their wanting to use it. The training also appeared to encourage fidelity in the context of a controlled trial leading one teacher to comment that, 'It's nice that this looks like a proper scientific research project, with the large sample size, randomised' (Teacher, School B). During the training session, teachers were given the opportunity to practice delivery of the revision materials, however, this was not taken up in all the training sessions that we observed.

In addition to the training, teachers were provided with a SMART Spaces handbook and access to a website with additional resources and guidance. Schools also received a support visit from the developer to follow up on the training and organisation prior to delivery. This training and support package as a whole was well received by the SMART leads with 79% agreeing or strongly agreeing with the statement, 'The SMART Spaces team provided everything we needed to run the intervention as a department.'

The handbook was rated highly: 72% of SMART leads and 59% of other teachers agreed or strongly agreed that it detailed everything they needed to know about SMART Spaces Revision. However, overall, 23% of all teachers did not respond to this item, which may be due to the fact that not all teachers reported receiving the handbook—20% of other teachers and 15% of SMART leads.

The website was accessed by a majority of teachers: 76% of lead teachers and 61% of other teachers. The website appears to have been more useful for lead teachers, 64% of whom strongly agreed or agreed that it was useful as opposed to 40% of other teachers. This may be to do with the lead teachers accessing updated resources from the website and disseminating them to colleagues.

An important aspect of the training was the support visit which took place after the training. The developers were able to organise support visits to 75% of schools and teachers were aware of these visits since 76% reported having had such a visit. Fifty-nine percent of SMART leads and 61% of other teachers agreed or strongly agreed that the support visit was useful in furthering delivery of SMART Spaces Revision, with only 4% and 3% disagreeing or strongly disagreeing with this statement. We note that there was a large proportion of missing responses (23%), which may reflect those reporting that they did not receive a visit. Cases study interviews and our communication with the developers suggest that this was sometimes one of the early SMART Spaces Revision intervention lessons, which account for only 45% of respondents reporting having rehearsed with another class before the intervention (see RQ12 below). Case study visits involved observation of developers giving feedback and subsequent interviews with teachers. Teachers valued the visits as they created an opportunity to practice prior to the intervention and provided feedback on the balance of elaboration and scripted delivery, as well as allowing teachers to ask questions—all of which supported the confidence of teachers in the delivery of the intervention.

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

The teachers were, in general, positive about the intervention and 70% of all teachers reported that they would be happy to deliver the intervention again in the future. This was slightly higher for heads of department (79%) than other teachers (66%), which may reflect higher 'buy-in' for the SMART leads. Further, 79% of SMART leads agreed or strongly agreed with the statement that 'overall, the department benefited from including SMART Spaces within revision practice'. However, while most of all teachers (66%) thought that SMART Spaces worked well for revision, teachers were more divided on whether the approach was a particularly engaging one for pupils: 40% neither agreed nor disagreed with the statement, 'The class seemed more motivated to learn during SMART Spaces than in normal lessons.'

A key aspect of the intervention is the set of slides that present the material that is to be revised in the lesson. We mapped each of the points in the AQA chemistry programme specification against the content of the slides used for the intervention. All the areas of subject content were represented within the slides, organised by how they appear in the examination papers. Particular attention was given to the language used so as to further support attainment by the use of precise terminology. Application of content knowledge was supported through the inclusion of worked examples, which vary over iterations of the SMART Spaces lessons. Teachers were positive about the slides for these reasons. A majority of teachers reported that they considered that these delivered the content well (Paper 1: 80%; Paper 2: 66%) and that they were a high quality revision resource (Paper 1: 74%; Paper 2: 70%). A further 17% neither agreed nor disagreed with this statement. The survey asked teachers how they thought revision through SMART Spaces Revision supports pupils in different aspects of their learning against the AQA combined science curriculum. Their responses are presented in Table 15 and

Table 16.

Table 15: SMART lead teachers responses to survey items around how SMART Spaces Revision supports pupils

SMART lead teachers only (n = 53) Revision through SMART Spaces supports pupils …	**Missing data**	**Disagree/ strongly disagree**	**Neither agree nor disagree**	**Agree/ strongly agree**
… to demonstrate knowledge and understanding of scientific ideas (AO1).	9%	4%	6%	81%
… to demonstrate knowledge and understanding of scientific processes, techniques, and procedures (AO1).	11%	6%	17%	66%
… in applying their knowledge and understanding of scientific ideas (AO2).	9%	40%	21%	30%
… in applying their knowledge and understanding of scientific processes, techniques, and procedures (AO2).	9%	42%	19%	30%
… to analyse, interpret, and evaluate scientific information, ideas, and evidence (AO3).	11%	60%	13%	15%

Table 16: Other teachers' responses to survey items around how SMART Spaces Revision supports pupils

Other teachers (n = 99) Revision through SMART Spaces supports pupils …	**Missing data**	**Disagree/ strongly disagree**	**Neither agree nor disagree**	**Agree/ strongly agree**
… to demonstrate knowledge and understanding of scientific ideas (AO1).	6%	1%	16%	77%
… to demonstrate knowledge and understanding of scientific processes, techniques, and procedures (AO1).	6%	5%	24%	65%
… in applying their knowledge and understanding of scientific ideas (AO2).	5%	23%	39%	32%

… in applying their knowledge and understanding of scientific processes, techniques, and procedures (AO2).	5%	24%	38%	32%
…. to analyse, interpret, and evaluate scientific information, ideas, and evidence (AO3).	5%	41%	36%	17%

It is clear that most teachers consider SMART Spaces Revision to be supportive of pupils demonstrating knowledge and understanding of scientific ideas and, to a slightly lesser extent, of scientific techniques and processes. These are both within Assessment Objective 1 (AO1) of the AQA combined science specification. For AO2, ‘application of knowledge and understanding’, teachers are divided with roughly a third of respondents agreeing or strongly agreeing, a third neither agreeing or disagreeing and a third disagreeing or strongly disagreeing. Marginally fewer heads of department are positive about support for AO2 than other teachers. For AO3, ‘analysis and interpretation of scientific information, ideas, and evidence’, a majority of teachers disagree or strongly disagree that SMART Spaces Revision is supportive. This is expected given the content of AO3 and AO2 as noted in the TIDieR framework.

The intervention materials cover areas of application, for example, the balancing of chemical equations; they are also designed to give different examples during later iterations so that application is supported. Our case study evidence leads us to suggest that teachers had pre-conceived framings that this kind of revision cannot support application however, and teachers were not convinced otherwise by the intervention. One teacher (in school E) commented that her class achieved good scores in AO2 (application) but sometimes struggled with recall—so SMART Spaces Revision was welcomed as part of revision—but she would still need to do examination practice with them.

While the majority of teachers were positive about SMART Spaces Revision, it is important to recognise that some were less enthusiastic and a small proportion had negative opinions. So while only 6% of all teachers reported that they would not be happy to teach the SMART Spaces lessons in the future, 15% disagreed or strongly disagreed with the statement, ‘The SMART Spaces lessons helped the class learn more easily than normal lessons.’ One potential reason for this may be that some teachers perceive it to be at odds with their preferred pedagogical approach. This was the case in two of the case study schools. For example, the head of department from School A does not feel that they will continue with SMART after the project because it does not fit with either her or the school’s pedagogical approach: ‘It’s not a method as a school that we will move forward with because if you look at Nuffield research, lecturing pupils just doesn’t work, but practice, practice, practice, testing does work.’

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

The majority of all teachers (84%) agreed or strongly agreed that they were confident in delivering SMART Spaces Revision, suggesting a high level of teacher engagement. This is supported by evidence from the pupil survey: 68% of pupils responding (n = 3694) agreed or strongly agreed that the teacher was confident at delivering the intervention. Our case study evidence suggests that, as predicted by the developers, this confidence grew through multiple deliveries. A teacher (School E) was typical in stating that she initially found delivery ‘very strange because I like that communication with them and just lecturing at them felt very out of the norm’. However, by the second and third run-throughs she was much more in her ‘comfort zone’. Observations from our case studies also suggest that pupils gained confidence in the second and third lessons, although teachers in two case study schools (A and E) reported that pupils lacked engagement and that pupils were ‘drifting off’ by the second and third run-through.

Given this generally high level of engagement, it is difficult to assess whether teacher engagement affected the quality of delivery. However, we have noted above that some teachers were less enthusiastic about the intervention. Our case study analysis suggests that some of these less enthusiastic teachers were willing to suspend their reservations about the efficacy of the approach and deliver with confidence. While survey responses suggest that, as would be expected, teachers were less confident in the impact of SMART Spaces Revision on attainment in AO2 and AO3 aspects of the curriculum (see RQ10, Table 15 and Table 16), our evaluation suggests that teacher professionalism limits any negative influence of doubts about the intervention on delivery. Teachers also make some adaptations in order to try and keep pupils engaged. An example of this was in School E, where a teacher identified pupil behaviour as a potential barrier as they had to be quiet and listen for longer than they are used to. He notes that some pupils are bored by the third run-through so changes the order of how the content is presented so as to support recall of the later sections of the presentation. A further example comes from School B, where a teacher felt that pupils 'were starting to drift away' even with the two 'breaks' (spacing activities). He also changed the order of slides so that some of the harder content comes earlier in the delivery in the second run-through. He deliberately slowed the delivery down for the third session. This is echoed by two other teachers in the same school (School B) who nevertheless reported that they were able to keep pupils engaged through behaviour management techniques and 'corralling' them. We saw one case (School A) where a teacher did not deliver the third run-through as she felt she had lost the engagement of the group. Her views of the intervention were broadly negative and it was reported by her and other teachers in the school that the approach did not fit the school's learning culture. The school's usual practice involves a lot of independent work by pupils and teachers considered that pupils did not respond well to this new approach, which they framed as 'lecturing'. Although a single case study, this does show how negative teacher and pupil views of the intervention might be interrelated.

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

Schools and teachers were encouraged (but not required) by the developer to rehearse teaching a SMART Spaces lesson with another class prior to the intervention. Although it features in the agreed logic model, at training this was not made a requirement, nor did it feature as a requirement in the handbook. Almost half of all teachers (45%) reported doing so while 51% reported not doing so.

As discussed in relation to RQ9, support visits by the developers tended to involve observation of early run-throughs of the scheduled SMART Spaces Revision intervention lessons to the trial classes, rather than additional intervention lessons. These did not provide further dosage of the intervention to the pupils, therefore. Nor did they provide teachers with additional dosage in the sense of practice runs of the intervention delivery.

Our interview protocol included a question around whether teachers practice the SMART Spaces lessons with another class and, in echo of the survey results, responses suggest that just less than a half did (heads of department concurred). Interview responses suggest that practice lessons took place with Year 10 classes that would benefit from the revision and that it was often a single SMART Spaces lesson rather than a sequence of three.

RQ12. (b) Do they adopt spaced learning in other chemistry revision lessons? [Dosage.]

We found no evidence to indicate that spaced learning was used or adapted in other chemistry revision lessons. Our case study evidence suggests that the SMART Spaces Revision intervention lessons constituted a large part, and in some cases all, of the chemistry revision practice immediately before the examinations for Paper 1 and Paper 2. In two cases study schools (Schools A and C), teachers reported that SMART Spaces had replaced all of the normal chemistry revision. In school B the SMART lead reported also doing examination practice beyond the SMART Spaces lessons (this did not come up at interview in the other three cases study schools). It is worth noting that the SMART Spaces Revision intervention was delivered at the end of the school year and it is possible that the approach might be more likely to influence teachers' approaches to chemistry revision in subsequent years after they have experienced SMART Spaces lessons for the first time.

RQ13. (a) To what extent do teachers adapt the materials and approach?

As we highlighted above (RQ6; fidelity), most teachers implemented SMART Spaces Revision without adapting the language or materials, aside from omitting slides that were beyond the examination content for foundation pupils. Lesson observation and interview evidence from the case studies suggests that teachers did sometimes adapt the delivery. For example, two teachers in School B reported slowing down delivery of some content or changing the level of detail or elaboration given. These changes all appeared to be within the spirit of the training, which promoted adaptation of delivery and materials for classes. Indeed, in training we saw the developers recommend that teachers include examples and references from previous teaching.

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

The majority of teachers reported that SMART Spaces Revision would be useful for other subjects beyond chemistry (79%, heads of department; 76%, other teachers). We found no evidence to indicate that schools had adapted their revision approaches in other subjects, although such adaptation might be more likely to occur in future years after teachers have experienced the SMART Spaces Revision approach.

RQ14. (a) Are all pupils responsive to SMART Spaces Revision and does it have reach: do all pupils perceive it to be an engaging and beneficial approach to revision? [Reach.]

Pupils appeared to be reasonably well engaged with the intervention: around half of pupils responding to the survey (46%) considered that SMART Space works well for revision and 49% considered that the approach would be useful for other subjects. A further 18% of pupils had no strong views. However, as indicated by the responses in Table 17, a minority of pupils reported negative perceptions about SMART Spaces Revision.

Table 17: Pupil survey responses around engagement with the intervention

All pupils (n = 3694)	Missing (blank)	Missing (unclear)	Disagree/ strongly disagree	Neither agree nor disagree	Agree/ strongly agree
The teacher was confident at delivering Spaced Learning	13%	5%	7%	9%	68%
I was enthusiastic to try Spaced Learning	14%	2%	18%	28%	39%
The Spaced Learning lessons helped me learn more easily than normal lessons	15%	1%	25%	24%	36%
I think Spaced Learning works well for revision	15%	1%	20%	18%	46%
The class as a whole enjoyed Spaced Learning	15%	1%	23%	28%	33%
I felt more motivated during Spaced Learning than in normal lessons	15%	1%	28%	25%	32%
I would be happy to try Spaced Learning again in future	15%	1%	23%	16%	45%
I think Spaced Learning would also be useful for other subjects	15%	2%	18%	16%	49%
I found the Spaced Learning lessons fun	15%	1%	20%	22%	41%
I found the Spaced Learning lessons helpful for revision	16%	1%	20%	19%	45%

These findings from the pupil survey correspond broadly with teacher perceptions of pupil engagement, as reported in relation to RQ11.

The impact analysis indicated no differences in the effect of the intervention on pupils with different levels of attainment (see Secondary Analysis and Table 12). However, around half of all teachers (48%) considered SMART Spaces Revision to be more useful for low attaining pupils. In contrast, only 27% of teachers considered SMART Spaces Revision to be less useful for high attaining pupils. Within case study interviews, the reasons given for considering SMART Spaces Revision to be more suitable for higher attaining pupils tended to follow the trend reported on in relation to RQ10 (see Table 15 and

Table 16), that teachers viewed it as more likely to support factual knowledge and recall. Higher attaining pupils were considered to require revision to focus on application and examination practice as they already have the content knowledge. This was framed by teachers as SMART Spaces Revision being ‘not challenging enough for them’ (teacher, School A) or a comment that it ‘doesn’t have the depth of content that they need’ (teacher, School D). Some teachers also felt that lower attaining pupils did not have the necessary ‘attention span’ for the focused content delivery and repetition (a different teacher in School A).

RQ14. (b) What contributes to pupil engagement (or disengagement)? [Reach]

We conducted interviews with small groups of pupils in four of the six case study schools. These suggest several themes that may affect pupil engagement, although we caution that the numbers of pupils involved were small and pupils had different (and at times contradictory) views.

Some pupils liked being presented with a summary and overview of the whole programme of study:

> *'It can cover the whole course in an hour and a half which, from the sounds of it, is brilliant ... it gives you the key points which you need to remember for the exams and then you can and fill in the gaps' (pupil, School E).*
>
> *'It helped me know what I do actually know and what I don't know so I can be more specific when I do, like, proper revision' (pupil, School E).*

On the other hand, some pupils felt that this overview placed too much emphasis on aspects of chemistry that they knew and not enough on the areas that they need to revise:

> *'It's not really going in because it's repetitive; it's stuff I already know … it would be better if it was less basic stuff' (pupil, School A).*
>
> *'Because it is just the basic facts, you feel like you are missing out on the more detailed stuff that you need to go through' (pupil, School E).*

In addition, some pupils reported liking the SMART Spaces Revision approach:

> *'It was good because in our classes it was interactive, so you could do it at the front with the teacher and every lesson more words would be removed and you would have to know more by yourself and even though it's in a class you are kind of working by yourself to figure out what you know and don't know' (pupil, school D).*

On the other hand, some pupils found the repetition unhelpful and a poor use of time:

> *'I found it more difficult to concentrate after the break … it was a whole hour and repetitive from yesterday' (pupil, School A).*
>
> *'I don't really engage with the information as much; I prefer to do exam practice this close to the exam … I feel I need to actually do something with the information. It's not very specific to what I know I need to revise … it's not the best use of time to focus on what I need to work on' (pupil, School D)*

These mixed views reflect pupils differing views on the efficacy of the SMART Spaces Revision approach. It is important to note that, as reported under RQ14(a) above, pupils appeared to be reasonably well engaged with the intervention.

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

Comparing revision practices for chemistry at home, we found no significant difference between control and treatment groups in the reported frequency of 'repeating practice of the same topics, with spaces in between'. This suggests that pupils in treatment schools are not employing spaced learning in their own revision any more than control pupils. Indeed, it is worth noting that almost all control group teachers (92%) reported encouraging their pupils to use this approach when revising at home, at least some of the time. However, pupils in the intervention group reported (practically significant) higher frequencies for two revision practices, including practising examination questions and working with a friend or family member. This suggests that, while there was no apparent increase in pupils' spacing practices, the intervention may support an increase in some revision practices, which may be due to an increased awareness of such practices.

In some intervention schools, pupils were supported in approaching revision through independent study. Discussion of this prompted one group of pupils to note in interview that:

> *'In normal revision you have quite free reign, you can get textbooks or read your CGP book [a commonly available commercial revision book] or you can do tests that they give you. But some people don't do this, so [SMART Spaces Revision is] helpful for people that don't really revise in the way they should be' (pupil, School D).*

We asked case study teachers and the groups of pupils whether being part of the intervention changed the way that pupils revised at home. Although in some schools the revision slides were made available, teachers did not think that spaced learning was being used at home, suggesting instead that revision notes and online tests were commonly used.

Usual practice

Usual practice was addressed in RQ16.

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

Many of the control group teachers reported using some form of spaced learning practices when teaching chemistry (62%) as well as in biology (52%) and physics (53%) at least some of the time.[15] This was broadly consistent with the control pupils' survey where 54% reported the use of spaced learning approaches during class revision in chemistry, 57% for biology, and 54% for physics. This may reflect an interest in spaced learning amongst schools who volunteered to take part in the study and it is noteworthy that a majority of the teachers (70%) reported using a related approach, interleaving, at least some of the time. However, spaced learning did not appear to be the main approach to revision: only 27% of control group teachers reported using spaced revision most or all the time in comparison to practice exam questions (96%), quizzes (51%), and dedicated websites (43%). For revision at home, teachers reported encouraging their pupils to use the following approaches most frequently: practising exam questions (72% report 'almost always') and use websites (42% report 'almost always') compared to 'repeatedly practice the same topics, with spaces in between' (20% report 'almost always').

Around two-fifths of control group teachers (43%) reported that they knew nothing or very little about spaced learning and only 30% reported that they knew a lot or a great deal about it. Hence, it is likely that in general teachers (and pupils) in control schools were using a much 'looser' definition and less systematic method of spaced learning than in the SMART Spaces Revision approach. Most control group teachers reported that support was available for revision in their school. Although only half (51%) of the teachers reported that their school had a revision policy, the majority of control teachers reported that their department provided GCSE revision materials (82%) and that they could seek support from colleagues in planning revision (92%).

A potentially surprising finding was that fewer pupils in the treatment group reported using spaced learning in biology or physics in comparison to the control group (for example, 44% of the intervention group pupils reported never using spaced learning approaches in biology compared to 15% in the control group) and these differences were statistically significant.[16] This should be treated with caution, because it may be that, as a result of their experiences in chemistry, many of the intervention group pupils conflated spaced learning with the specific SMART Spaces Revision approach.

The SMART Spaces Revision logic model

RQ17. To what extent does the logic model (Figure 3 and Figure 4) adequately describe the mechanism by which the SMART Spaces Revision intervention effected change (if any)? [Differentiation, intervention characteristics.]

The IPE results reported above indicate a high degree of fidelity to the approach as described in the TIDieR framework and logic model. Aside from some relatively minor differences (for example, some teachers not practising SMART

---

[15] Percentages refer to 'some', 'most', or 'almost always' responses combined.

[16] Intervention report of spaced learning in Biology revision (Mean=1.56, SD=0.82); control report (Mean=2.30, SD=0.92) was significantly greater, $t(5955.6)=34.2$, $p<0.01$.

Spaces Revision prior to the intervention), teachers and schools largely implemented the intervention as envisaged by the developer. Hence, the null result of the intervention is unlikely to be the result of poor implementation.

It was hypothesised in the logic model that pupil and teacher engagement would have a moderating effect on the learning and teaching of SMART Spaces Revision, respectively. We found no evidence for an effect of pupil engagement as reported in the impact evaluation results above. As noted above (RQ11), it was difficult to assess whether teacher engagement affected the quality of delivery as teacher engagement was relatively high. However, there was some evidence to suggest that less enthusiastic teachers were prepared to put aside their reservations about the approach and deliver the lessons as envisaged by the developer.

Two areas of the logic model which might be further developed were this intervention is to be taken to a larger scale trial. First, 'school implementation factors' and 'school approach to science revision' are clearly significant, with potential barriers being the integration of SMART Spaces lessons within the school calendar and accommodating longer lessons where they are not normally one hour long. As such, the consideration of how to take account of such issues within the intervention could be furthered if the intervention were developed further. We also suggest that integration with other forms of revision (for example, examination practice) could be supported in guidance and support. The second area concerns pupil self-revision and the use of slides and spacing. We found no evidence that resources or the approach from SMART Spaces Revision were adopted by pupils (nor evidence that it was not). Consideration might be given to the role of this within the intervention and whether it should be supported further or omitted from the logic model as being outside the scope of the intervention.

Compliance

As noted above under the compliance analysis in the impact analysis, levels of compliance by schools were high. The IPE analysis above provides further evidence of high compliance. This further suggests that the intervention was relatively straightforward to implement at both school and classroom levels.

Cost

Delivery of the SMART Spaces Revision intervention cost approximately £3,267 per school. The majority of costs are incurred during the first year and are associated with the management and delivery of the training and in-school coaching, manuals and website resources, and spacing activity resources. There may be some additional resources associated with training new staff due to staff turnover, although the expectation is that this would be carried out by existing teachers cascading the approach and, hence, the cost would be in teacher time.

To calculate the total cost per pupil over three years, we assumed that the intervention would be delivered to each Year 11 cohort of pupils. We assumed 120 pupils per school, as this was the average number of pupils per intervention school at allocation. Based on these assumptions, the total cost per pupil per school over three years is £9.07 (set out in Table 18). The cumulative cost breakdown is set out in Table 19 (figures rounded to the nearest pound).

Evidence from case study schools did not indicate any additional school costs or time requirements. In particular, aside from the time required to timetable the lesson blocks during GCSE revision weeks, the intervention required little, or no, additional preparation time for schools or teachers.

Table 18: Cost of delivering the SMART Spaces Revision intervention

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
Developer: management of training and support; ongoing schools support; quality control	Start Up (Y1 only)	£960.28	£960.28	£2.67
Initial training (trainer costs)	Start Up (Y1 only)	£411.11	£411.11	£1.14
Coaching visits (trainer costs and scheduling staff time)	Start Up (Y1 only)	£756.48	£756.48	£2.10
Trainer travel for initial training and follow-up visits	Start Up (Y1 only)	£824.74	£824.74	£2.29
Manuals for teachers and website resources	Start Up (Y1 only)	£270.28	£270.28	£0.75
Spacing activity equipment (juggling balls, balloons): initial provision	Start Up (Y1 only)	£26.46	£26.46	£0.07
Spacing activity equipment (juggling balls, balloons): replacement	Recurring	£8.82	£17.64	£0.05
Total			£3,266.99	£3,266.99/120/3= £9.07

Table 19: Cumulative costs of the SMART Spaces Revision intervention—assuming delivery over three years

	Year 1	Year 2	Year 3
SMART Spaces	£3,249	£9	£9

Conclusion

Table 20: Key conclusions

Key conclusions
1. Pupils in the SMART Spaces Revision schools, on average, made no additional months' progress in chemistry attainment in AQA Combined Science GCSE compared to pupils in the control schools. This result had a high security rating.
2. Pupils in the SMART Spaces Revision schools, on average, made no additional months' progress in overall GCSE Combined Award Science compared to pupils in the control schools.
3. Pupils eligible for free school meals (FSM) in the SMART Spaces Revision schools, on average, made no additional months' progress in chemistry attainment at AQA Combined Science GCSE compared to FSM-eligible pupils in the control schools.
4. The SMART Spaces Revision theory of change predicts that spaced learning primarily improves recall of knowledge. There was weak evidence that intervention pupils had improved chemistry knowledge attainment compared to the control pupils. However, the effect was small and did not meet the threshold set by the evaluator that it was non-zero. Also, it may be at the expense of negative impact on chemistry application and analysis attainment.
5. The SMART Spaces Revision intervention was delivered as intended in schools, indicated through observations, pupil and teacher surveys, and interviews. Teachers showed a high level of support for the intervention based on teacher surveys and interviews, with 70% reporting that they would be happy to deliver SMART Spaces Revision in the future.

Impact evaluation and IPE integration

Evidence to support the logic model

In general, the intervention was delivered as described in the logic model. We found compliance and fidelity to be high: schools and teachers largely implemented the intervention as described in the TIDieR framework and fidelity to the SMART teaching approach appeared to be reasonably high. Two thirds of classes received the full set of SMART Spaces lessons and 84% received at least four lessons. Moreover, the majority (around 90%) of these lessons were taught as intended.

Survey and case study evidence indicated that implementation was largely relatively straightforward. There were some challenges around timetabling and fitting SMART sessions into shorter lesson periods. Most schools were able to resolve these issues. In this regard, the role of the SMART lead teacher appeared to be important in 'problem-solving' and in ensuring the SMART lessons took place.

The logic model hypothesised that teacher engagement (enthusiasm, confidence, and buy-in) would influence delivery of SMART Spaces Revision and, therefore, influence pupil outcomes. Teachers were largely supportive of the intervention. Most teachers taught the full set of six SMART lessons and reported that they delivered the materials largely as intended. The training, SMART coaching visit, handbook, and materials were all viewed positively by most teachers. Case study evidence suggested that, as hypothesised, teacher confidence grew through multiple deliveries of the intervention. There were some teachers who were less enthusiastic about SMART Spaces Revision and a small proportion were negative about the intervention;: some 15% of teachers disagreed or strongly disagreed with the statement, 'The SMART Spaces lessons helped the class learn more easily than normal lessons.' The case study data suggested that one potential reason for this may be that some teachers perceive the programme to be at odds with their, and their schools', preferred pedagogical approaches. However, the case study data suggested that teacher engagement was not as important an implementation factor as hypothesised in the logic model. Indeed, we found evidence that that many of the less enthusiastic teachers were prepared to put aside their reservations about the approach and deliver the lessons as envisaged in the TIDieR framework.

The logic model hypothesised that pupil engagement would influence pupil outcomes. In general, pupils appeared to be reasonably well engaged with the intervention: around half of pupils responding to the survey considered that SMART Spaces worked well for chemistry revision, although a sizeable minority (20%) of pupils reported negative attitudes towards the intervention. However, we found no evidence for a moderating effect of pupil engagement on the effect as hypothesised in the logic model: more engaged pupils were not more likely to gain from the intervention. Survey evidence from pupils indicated that pupils did not adopt spacing practice within their own self-revision practices as hypothesised in the logic model, although the timing of delivery, just prior to the GCSE examination, is not likely to have provided sufficient time to enable pupils to adopt these practices for these examinations.

The SMART Spaces Revision intervention is directed at all pupils entered for the GCSE Combined Science Award and thus excludes the lowest and highest attainment pupils in the Year 11 cohort. Nevertheless, the Combined Science GCSE accounts for the majority of entries in science and there remain differences in attainment across these entries. Despite this, the SMART Spaces Revision intervention's only differentiation is in the content covered for classes entered for either the foundation or the higher tiers; there is no other differentiation in terms of pedagogy or expectations. Around half of all teachers (48%) considered SMART Spaces Revision to be more useful for low attaining pupils while just over a quarter (27%) considered it to be less useful for high attaining pupils. However, the impact analysis indicated no differences in the effect of the intervention on pupils with different levels of attainment.

Interpretation

The impact evaluation results for all outcomes indicated effects that were all very small in magnitude, indicating either zero month's progress or less in all outcomes aside from AO1 (chemistry knowledge) for which there was an effect equivalent to one month's progress. As discussed above, the intervention was delivered as described in the logic model with reasonably good compliance and reasonably high fidelity. Hence, the null result of the intervention on the primary outcome (chemistry attainment) is unlikely to be due to poor implementation.

The logic model hypothesised that the intervention would improve factual recall, which would in turn have an impact on all three areas of assessment in chemistry: knowledge (AO1), application (AO2), and analysis (AO3). Since AO1 is focused on factual recall, the impact on AO1 was hypothesised to be greater than for AO2 and AO3. Exploratory analysis of pupil attainment in each of these assessment areas suggests that pupils in SMART Spaces schools made on average a gain of $d = 0.06$ in knowledge (AO1), which is equivalent to one months' progress. However, this result was supported by at best weak evidence and the statistical evidence did not meet the threshold set by the evaluator to conclude that the true impact was non-zero. It may also be at the expense of poor progress in AO2 and AO3, which both show negative effects of a similar magnitude.

Survey evidence from teachers and pupils indicated that around half of the control schools used some aspects of spaced learning in both teaching and revision. This may reflect an interest in spaced learning amongst schools who volunteered to take part in the trial. However, in most of these schools, the approach to spaced learning seems unlikely to be as systematic as SMART Spaces Revision since around two fifths (43%) of control group teachers reported that they knew very little or nothing about spaced learning. Hence, the null result is unlikely to be due to 'leakage' of the intervention, that is, control schools implementing aspects of spaced learning.

The impact evaluation results also showed no difference between subgroups: pupils in receipt of free school meals, girls compared to boys, or low compared to high attainment. On the other hand, teachers were, in general, positive about the intervention and it did not appear to deliver worse outcomes than other revision approaches. Moreover, the materials provide a complete and coherent summary of the CGSE combined award chemistry content, although the content is skewed more towards AO1 (knowledge) than either AO2 (application) or AO3 (analysis). Hence, since the intervention appears to be relatively easy to implement, there may be some value in suggesting the approach to schools as one element to consider when constructing their GCSE chemistry revision programme.

Limitations and lessons learned

There are several limitations to this study. First, SMART Spaces Revision is focused on revision immediately prior to a GCSE and we have not investigated the wider effects of spaced learning in this trial. We did conduct a small pilot study focused on SMART Spaces as a teaching approach, published in a separate report, which produced somewhat inconclusive results (Hodgen et al., 2023). Second, all participating schools (control and intervention) volunteered to take part in this trial and some control schools did use some aspects of spacing in teaching and revision. Third, the trial involved only schools that entered pupils for the AQA Combined Science GCSE award and, hence, did not involve the full range of pupil attainment in the cohort. However, we do not consider the focus on one awarding body, AQA, to be a serious limitation.

One positive aspect of this study concerns the ease of implementation. Although there were some challenges around timetabling, most schools were able to resolve these issues and found implementation relatively straightforward. In addition, teacher engagement was high and even teachers who were sceptical of the approach appeared, in general, willing to implement the approach. The IPE suggests two main reasons for this. First, the materials and training are comprehensive and coherent and were delivered consistently across intervention schools. Second, SMART Spaces Revision is a ‘contained’ intervention: delivery takes place in a lesson period and does not require more general changes to teachers’ or schools’ practices.

In terms of specific feedback for the SMART Spaces Revision intervention, the IPE indicated that schools may require more guidance on timetabling and on fitting lessons into shorter lesson slots. The IPE also indicated that guidance is needed on the integration of spaced approaches with other forms of revision and on ways of engaging pupils in aspects of spaced revision.

Future research and publications

As indicated above, it is important to emphasise that although we found no positive effects on chemistry or science attainment, we also found no evidence to indicate that SMART Spaces Revision had a negative effect on attainment. Indeed, this trial provides evidence that the SMART Spaces Revision intervention could be a useful element within a school’s revision programme. It is also important to distinguish SMART Spaces Revision from spaced learning and revision involving spacing more generally. This trial should not be interpreted as providing evidence for or against the efficacy of spaced learning more generally. Rather, the results indicate that this particular approach, SMART Spaces Revision, did not have an additional impact on pupil outcomes when compared to usual practice. There remains a need for research focused on the design, implementation, and efficacy of spaced learning teaching approaches and interventions. The current trial provides some evidence that teachers are willing to implement such approaches, but the challenge of translating the broad research findings about learning involving spacing into actionable teaching strategies and interventions remains.

We plan to publish several academic journal papers summarising and extending the findings of this evaluation report as well as discussing the implications for future research on spaced learning and the design of interventions more generally.

References

Bell, M. C., Kawadri, N., Simone, P. M. and Wiseheart, M. (2014) 'Long-term Memory, Sleep, and the Spacing Effect', *Memory*, 22 (3), pp. 276–83. http://doi.org/10.1080/09658211.2013.778294

Benton, T. and Sutch, T. (2014) 'Analysis of Use of Key Stage 2 Data in GCSE Predictions', report commissioned by Ofqual ref:14/5471, Coventry: Office of Qualifications and Examinations Regulation.

Braun, V. and Clarke, V. (2006) 'Using Thematic Analysis in Psychology', *Qualitative Research in Psychology,* 3 (2), pp. 77–101. doi:10.1191/1478088706qp063oa

Cepeda, N. J. , Pashler, H., Vul, E., Wixted, J. T. and Rohrer, D. (2006) 'Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis', *Psychological Bulletin*, 132 (3), pp. 354–380.

Dessing, J. C., Rey, F. P. and Beek, P. J. (2012) 'Gaze Fixation Improves the Stability of Expert Juggling', *Experimental Brain Research*, 216 (4), pp. 635–644. http://doi.org/10.1007/s00221-011-2967-6

Donovan, J. J. and Radosevich, D. J. (1999) 'A Meta-Analytic Review of the Distribution of Practice Effect: Now You See It, Now You Don't', *Applied Psychology*, 84 (5), pp. 795–805. https://doi.org/10.1037/0021-9010.84.5.795

Ebbinghaus, H. (1913) *Memory: A Contribution to Experimental Psychology*, Teachers College, Columbia University.

EEF (2015) 'Intra-Cluster Correlation Coefficients', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/ICC_2015.pdf

EEF (2018) 'Statistical Analysis Guidance', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

Gaure, S., McDermott, G., Vargas, M., Werner, K. D., Stigler, M. and Lüdecke, D. (2022) 'lfe: Linear Group Fixed Effects'. https://cran.r-project.org/web/packages/lfe/index.html

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

Gelman, A. and Rubin, D. B. (1992) 'Inference from Iterative Simulation Using Multiple Sequences', *Statistical Science*, 7 (4), pp. 457–72.

Hodgen, J., Anders, J., Bretscher, N. and Hardman, M. (2023) 'SMART Spaces: Spaced Learning (Teaching Pilot Trial) Evaluation Protocol', London: Education Endowment Foundation.

Humphrey, N., Ledrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook', London: Education Endowment Foundation.

Kruschke, J. K. and Liddell, T. M. (2018) 'The Bayesian New Statistics: Hypothesis Testing, Estimation, Meta-Analysis, and Power Analysis from a Bayesian Perspective', *Psychonomic Bulletin and Review*, 25 (1), pp. 178–206. doi:10.3758/s13423-016-1221-4

McDonald, J. H. (2014) *Handbook of Biological Statistics (3rd edn)*, Baltimore, ML: Sparky House.

Miles, S. W. (2014) 'Spaced vs. Massed Distribution Instruction for L2 Grammar Learning', *System*, 42, pp. 412–428. http://dx.doi.org/10.1016/j.system.2014.01.014

Nieuwenhuis, S., Elzinga, B. M., Ras, P. H., Berends, F., Duijs, P., Samara, Z. and Slagter, H. A. (2013) 'Bilateral Saccadic Eye Movements and Tactile Stimulation, But Not Auditory Stimulation, Enhance Memory Retrieval', *Brain and Cognition*, 81(1), pp. 52–56. http://doi.org/10.1016/j.bandc.2012.10.003

Nitz, S., Prechtl, H. and Nerdel, C. (2014) 'Survey of Classroom Use of Representations: Development, Field Test and Multilevel Analysis', *Learning Environments Research*, 17 (3), pp. 401–422. doi:10.1007/s10984-014-9166-x

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

Perry, T., Lea, R., Jørgensen, C. R., Cordingley, P., Shapiro, K. and Youdell, D. (2021) 'Cognitive Science in the Classroom: Evidence and Practice Review', 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. and Kirk, I. J. (2015) 'Influence of Physical Activity on Human Sensory Long-Term Potentiation', *Journal of the International Neuropsychological Society*, 21 (10), pp. 831–840. http://doi.org/10.1017/S1355617715001095

Smolen, P., Zhang, Y. and 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

Stern, P. (2009) 'Making Memories Last', *Science Signaling*, 2(85). http://doi.org/10.1126/scisignal.2009ec285

Taylor, K. and Rohrer, D. (2010) 'The Effects of Interleaved Practice', *Applied Cognitive Psychology*, 24 (6), pp. 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. and 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. and Knecht, S. (2007) 'High Impact Running Improves Learning', *Neurobiology of Learning and Memory*, 87 (4), pp. 597–609. https://doi.org/10.1016/j.nlm.2006.11.00

Appendix A: EEF cost rating

Figure 10: Cost rating

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

OUTCOME: Chemistry subscale of GCSE combined science

Rating	Criteria for rating			Initial score	Adjust	Final score
	Design	**MDES**	**Attrition**			
5	Randomised design	≤ 0.2	0–10%			
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21–0.29	11–20%	4		4
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30–0.39	21–30%		Adjustment for threats to internal validity **[0]**	
2	Design for comparison that considers selection only on some relevant confounders	0.40–0.49	31–40%			
1	Design for comparison that does not consider selection on any relevant confounders	0.50–0.59	41–50%			
0	No comparator	≥ 0.6	>50%			

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	No significant risk apparent.
Threat 2: Concurrent interventions	Low	Survey findings suggest that some aspects of spaced learning were used by teachers in control schools. However, their approach to spaced learning seemed unlikely to be as systematic as SMART spaces and it did not appear to be a focus of 'business as usual' revision practice.
Threat 3: Experimental effects	Low	Randomisation was undertaken at the school level minimising risk of contamination. IPE suggests there were no important instances of compensatory rivalry.
Threat 4: Implementation fidelity	Low	Intervention was delivered as described in the logic model with high compliance and good fidelity.
Threat 5: Missing data	Moderate	Attrition of 15.1% with some evidence of differential attrition. However, analyses accounting for missing data are similar to the complete-cases analysis.
Threat 6: Measurement of outcomes	Low	Outcome measures are appropriate and appear to have good statistical properties.
Threat 7: Selective reporting	Low	Reporting is appropriate. The study was registered, the analytical approach was pre-specified.

- **Initial padlock score:** 4 padlocks
- **Reason for adjustment for threats to validity**: N/A
- **Final padlock score:** 4 Padlocks

Further appendices

Please see accompanying document '*Further appendices*'

Education Endowment Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn