



## OPEN ACCESS

## EDITED BY

Aeen Mohammadi,  
Tehran University of Medical Sciences, Iran

## REVIEWED BY

Maryam Aalaa,  
Tehran University of Medical Sciences, Iran  
Snor Bayazidi,  
Tehran University of Medical Sciences, Iran

## \*CORRESPONDENCE

Liam O'Hare  
✉ lohare@qub.ac.uk

RECEIVED 03 April 2023

ACCEPTED 23 August 2023

PUBLISHED 19 September 2023

## CITATION

O'Hare L, Stark P, Gittner A, McGuinness C,  
Thurston A and Biggart A (2023) A pilot  
randomized controlled trial comparing the  
effectiveness of different spaced learning  
models used during school examination  
revision: the *SMART Spaces 24/10* model.  
*Front. Educ.* 8:1199617.  
doi: 10.3389/educ.2023.1199617

## COPYRIGHT

© 2023 O'Hare, Stark, Gittner, McGuinness,  
Thurston and Biggart. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).  
The use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in this  
journal is cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# A pilot randomized controlled trial comparing the effectiveness of different spaced learning models used during school examination revision: the *SMART Spaces 24/10* model

Liam O'Hare<sup>1\*</sup>, Patrick Stark<sup>2</sup>, Alastair Gittner<sup>3</sup>,  
Carol McGuinness<sup>1</sup>, Allen Thurston<sup>1</sup> and Andy Biggart<sup>1</sup>

<sup>1</sup>School of Social Sciences, Education and Social Work, Queen's University Belfast, Belfast, United Kingdom, <sup>2</sup>School of Nursing and Midwifery, Queen's University Belfast, Belfast, United Kingdom, <sup>3</sup>Hallam Teaching School Alliance, Sheffield, United Kingdom

**Introduction:** The main aim of this pilot study was to compare the efficacy of using different spaced learning models during school examination revision on pupil attainment. Spaced learning is using intervals between periods of learning rather than learning content all at one time.

**Methods:** Three spaced learning models with different inter-study intervals (ISI's) were co-designed by teachers and researchers using research evidence and practice knowledge. A pilot randomized controlled trial compared the three ISI models against control groups in 12 UK secondary schools' science classes (pupil  $n = 408$ ). The effects on attainment of each model were assessed using pre and post-tests of science attainment.

**Results:** The results showed that all three models were feasible for use in a classroom. The spacing model using ISI of 24-h spaces between and 10-min spaces within revision sessions was the only significant one for improving attainment against a control group (effect size  $d = 0.19$ ,  $p = <0.05$ ). The study also found that student engagement with the spaced learning program was a statistically significant predictor of increased pupil attainment.

**Discussion:** The study demonstrates the potential benefits of applying spaced learning to exam revision, with the most optimal ISI model found to be the *SMART Spaces 24/10* model.

## KEYWORDS

spaced learning, inter-study interval, exam revision, attainment, randomized control trial

## Introduction

Spaced learning is a process in which periods of learning are separated in time by an inter-study interval (ISI). An ISI is the length of time between the periods of learning. The ISI may be as brief as a few seconds, or as long as weeks and months. When the efficacy of a spaced learning strategy is examined, it is often compared with a "massed learning" approach. Massed

learning is where all the to-be-learned content occurs in one constant block, without any intervals between study periods.

Spaced learning is often discussed in education as the result of a distributed practice or the “spacing effect” (Chen et al., 2018; Wiseheart et al., 2019). The “spacing effect” refers to the benefit of spaced learning has on memory and retention of information over massed learning (Ebbinghaus, 1885). The spacing effect is a well-researched and remarkably stable effect in psychological (Zwaan et al., 2017) and educational research (Dunlosky et al., 2013). This is a particularly pertinent point given the recent debates around the replicability of psychological phenomenon (Motyl et al., 2017). On this basis, because of its robustness, the spacing effect has great potential to be applied in educational settings (Firth, 2021) to improve attainment through its use as a theory of intervention, i.e., a theory within an evidence-based program that explains how program activities change program participant outcomes (Connolly et al., 2017 - Chapter 2).

## Inter-study intervals and attainment: perspectives, theory and evidence

The robustness and replication of the spacing effect on memory, retention and attainment has been supported through many literature reviews over an extended period (Connolly et al., 2017). The following paragraphs outline the main findings from some of these reviews. However, there is a particular focus on the effectiveness of different ISIs on retention performance and attainment as this is the main outcome variable compared in this current study.

The optimal length of ISIs depend on the required retention interval (RI). Moss (1995) reviewed 120 articles on the spacing effect, comparing various types of learning material (verbal information, intellectual skill, and motor learning). The review found longer spacing intervals improved the learning of verbal information and motor skills in over 80% of the studies reviewed. Donovan and Radosevich's (1999) review also found stronger spacing effects by increasing the ISI (mean ES for spacing = 0.46). Janiszewski et al. (2003) reviewed 97 articles on spacing effects and space lengths for various types of tasks. Again, the largest spacing effects arose from longer spaces (mean ES = 0.57). Cepeda et al. (2006) reviewed 317 spacing effect experiments in 184 articles, including children, adults and older adults as learners. All but 12 of the 317 studies showed a benefit of spaced learning over massed study. Cepeda et al. found that increasing the spacing interval increased recall, but too long a spacing interval for a given retention period reduced recall. Cepeda et al. determined that the spacing interval during studying should increase as the retention interval increases to optimize recall. When the retention interval was less than 1 min, the optimal space was also less than 1 min. At least a one-month space was necessary for optimal recall after 6 months. What activity is undertaken during the space has also been found to be important for learning and attainment. Specifically, sleep during the activity may be crucial. Bell et al. (2014) found that sleep during 12-h spaces between periods of learning Swahili-English pairs led to better retention performance than the same length of space with no sleep.

Studies within the field of neuroscience have explored the neurobiological basis of the spacing effect, but with much shorter ISIs. These studies, which have largely been conducted with animals, have focused on recording neurobiological indicators of a precursor to memory formation—“long term potentiation” (LTP)—a process through which synapses may become stronger after being stimulated, and thus

transmit a long-lasting signal between neurons. Mauelshagen et al. (1998) exposed synapses removed from *Aplysia* (marine mollusks) to serotonin in either five bursts of 5 min with 15-min intervals, or one long massed exposure of 25 min. They found that electrical responses to stimulation 24h later (which is representative of the type of cell activity important for long-term memory formation) was greater for spaced stimulation than massed stimulation. Fields (2009) reported that rat synapses produced increased levels of protein and gene markers of memory formation (CREB and zif268) and twice the voltage of electrical activity after being stimulated in three bursts with 10-min spaces than when they were stimulated in a massed pattern. The neuroscience literature also points to the benefit of increasing the space length for better retention. Kramar et al. (2012) found twice as much LTP in rat brain cells when stimulating in 60-min spaced intervals rather than 30- or 10-min intervals. Zhang et al. (2012) found that a spacing protocol including a 30-min space in the stimulation of mollusk brain cells led to greater activity than 15-min spaces. The earliest neuroscience evidence of spacing effects in humans was gathered using experimental paradigms that did not include explicit spaced learning tasks but did include a comparison of spaced and massed repetitions. Van Strien et al. (2007) found a larger change in event-related potentials (electrical recording of brain events) associated with memory search and template matching (N400 and LPC) in response to massed rather than spaced presentations learning of repeated words. This suggests that massed presentation of learning resulted in more difficulty performing these two aspects of recall.

To date, we are aware of only two studies involving an explicit spaced learning task and the recording of human brain activity. Mollison and Curran (2014) compared the paired learning of nouns and pictures from two repetitions presented in either a massed or spaced format (12-s ISIs) and found ERP event-related potential evidence of repetition suppression (a reduced response to material when presented repeatedly) for massed but not spaced presentations. Furthermore, the spaced items were remembered with higher accuracy than the massed items. This result may indicate that attention to repeated items is better when a spacing strategy is used. Functional magnetic resonance imaging (fMRI) has also been used to explore spaced learning. Xue et al. (2011) found more activity in a brain area associated with face recognition (bilateral fusiform gyrus) for spaced learning of novel faces than for massed learning.

The earlier cognitive psychology reviews of spaced learning literature highlight robust spacing effects mainly for simple tasks, but there is educational practice literature on the spacing effect for complex learning more like everyday learning that would occur in classrooms and its impact on retention performance. For example, the attainment benefits of spaced learning has been demonstrated in a wide range of educational contexts, e.g., vocabulary retention in elementary schools (Sobel et al., 2011), mathematics attainment in secondary school (Barzagar Nazari and Ebersbach, 2019) attainment in English as a foreign language (Namaziandost et al., 2020) and examination performance for psychology undergraduates (Gurung and Burns, 2019). Like the lab-based literature, the length of ISI is a key consideration in educational practice literature. Miles (2014) found that students learning English as a second language using a spacing protocol of 1-week and 4-weeks scored more highly on a subsequent language task than students using massed learning. The retention interval was 5 weeks. Another study using a long retention interval of 5 weeks found that psychology undergraduates in Canada taught using a spaced protocol with eight-day ISIs showed better

retention after 5 weeks than when 24 h ISIs were used (Kapler et al., 2015). This fits with the Cepeda et al. (2006) finding that 24-h ISIs are optimal for up to a 28-day retention interval. Retention for 5 weeks, using a complex task, needs more than a 24-h interval. Bird (2011) compared ISIs of three and 14 days for spaced learning lessons of English grammar for university students learning English. No difference was found after a retention interval of 7 days, but at 60 days, the shorter space group had decreased in accuracy and the longer space group's score remained consistent with their 7-day score. This study suggests that ISIs of 3 days or 14 days may be too long to see a benefit for 7 days' retention. Regarding very long retention intervals, Carpenter et al. (2009) found that children using an ISI of 16 weeks recalled more history facts over a retention interval of 9 months than children using an ISI of 1 week.

## Optimizing spaced learning for school exam revision

For better or worse, the key success outcome of schooling is pupils' performance on national examinations (Friedman and Laurison, 2020). Therefore, the need for pupils and students to revise and prepare for high stakes examinations has been, and continues to be, a substantial focus for schools around the world. All the research evidence on spaced learning is useful but is constrained by the context in which it is applied. There are several contextual considerations when designing a spaced learning program for examination revision purposes.

Head teachers and classroom teachers are keenly aware of the need to ensure that all students are successful regardless of their socio-economic background. Consequently, disadvantaged students are a particular focus for schools. It is however, recognized by teachers that study skills may be less well developed for disadvantaged pupils (Putwain, 2008). For some students the only examination revision they may complete will be in school as their home environment might not be conducive to focussed study. Relatedly, there is some evidence that homework fails to provide an advantage for disadvantaged pupils (Rønning, 2011).

It is important to consider current revision practices in school to assess what a spaced learning exam revision program might substitute. Current techniques employed by schools in the UK for GCSE science include the use of past papers, through the availability of previous GCSE science papers from examination boards. This technique involves pupils sitting previous exam papers as a practice test. The use of practice testing has an evidence base of similar longevity as the spacing effect (Abott, 1909) and the practice testing effect too is robust (see Rawson and Dunlosky, 2011 for review). Carpenter et al. (2009) suggested that this technique is effective through the triggering of elaborative retrieval processes. While much of the literature on practice testing has been on verbal attainment tasks, such as paired associate learning and word lists, there is an increasing evidence base for benefits for attainment on more complex tasks such as multiplication facts, word definitions, science facts and key term concepts (Dunlosky et al., 2013). However, there is one caveat. For practice testing to be most effective, feedback must be given (Dunlosky et al., 2013), and this introduces the major barrier for the feasible use of practice papers—marking and feedback. The initial practice time per student may not be high, but to make the technique effective the

required marking of past papers is hugely demanding in an educational setting.

Another consideration regarding using the spacing effect within educational practice and exam revision is pupil engagement. A wide variety of perspectives suggest the issues of student engagement in spaced learning is worthy of consideration. Generally, participant engagement or responsiveness is offered as an important factor for implementing interventions with fidelity and impact (O'Hare, 2014; Connolly et al., 2017; O'Hare et al., 2017, 2018). In addition, the idea of “chunking,” breaking down learning into short episodes and changes in activity by teachers are understood to assist student attention (Gobet, 2005). There is also some logic in the notion that science teachers might particularly engage with a spaced learning approach to exam revision as it is underpinned by substantial scientific evidence.

## Study rationale

In designing the current study, the authors argue that the overwhelming evidence for the spacing effect moves it beyond a theoretical position and closer to a replicable stable effect, both in laboratory and classroom settings. Building on these solid foundations the educational research questions now turn to how best to optimize this effect for specific educational settings and outcomes, e.g., improving attainment in a real-world science classroom. Thus, this study synthesizes the theories and evidence from cognitive psychology, neuroscience and educational practice literature, to develop a range of Inter Study Intervals (ISI) to check for their feasible and engaging use by science teachers for revision purposes in science classrooms with the goal of improving pupil science attainment. The study also compares these ISI's against each other in an RCT design to identify the optimum ISI model for future application and investigation.

The cognitive psychology theory and evidence on spacing suggest that the ISI should be at least 1 day for the length of retention interval that is desirable for exam preparation (weeks and months). The cognitive psychology evidence also shows that spaces of weeks or months are advantageous for long-term retention performance. However, one aim of the current research, is to apply spaced learning to science revision in schools in a way that is feasible in terms of the practicalities of school classrooms and schedules. Thus 24-h spaces are chosen as one potential spacing strategy for an applied spaced learning program. For the purposes of revision in the lead-up to examinations, the use of an ISI of 24 h also potentially facilitates the benefits of sleep for memory formation (Bell et al., 2014). It also avoids the demands of multiple lessons within the one school day which is not unfeasible for school schedules, especially in high school education. The 24-h space is also likely to be the most appropriate ISI for mid-length retention, considering Cepeda et al.'s findings of too long a spacing protocol being detrimental, as the prioritized outcome of a revision program is exam performance, close in time to receiving the program, and not long-term retention. Cepeda et al. found a 24-h space to be the most advantageous for retention intervals of between 2 and 28 days, which is a realistic interval of time for delivering a revision program in schools.

Despite the justified concern about the difficulty and validity of translating neuroscience evidence in the classroom (Donoghue and Horvath, 2016; Horvath and Donoghue, 2016) the neuroscience literature did inspire some classroom applications of spaced learning and these may have practical applications. Kelley and Watson (2013)

adopted Fields (2009) spacing protocol from his neuroscience work and successfully employed spaced learning strategies in the classroom with children aged 14–15 years in England, in a quasi-experimental study. They claimed that 90 min of spaced learning (three periods of 20 min of teaching, interspersed with 10-min intervals of distractor activities) produced retention of the information that was not significantly different to 4 months of typical teaching (massed learning), despite significantly less teaching time. However, Timmer et al. (2020) found no significant benefit of using short 5-min spaces with a single lecture for medical students. The students did, however, give positive feedback on the lessons and the Timmer et al. (2020) called for more research into optimum spacing patterns. The neuroscience theory and evidence indicates that short spaces (60 min or less) can still reveal a spacing effect, even if longer spaces may have an advantage in retention length. The 10-min ISI has inherent value in the classroom due to the length of lessons. Considering the intense, rapid delivery style of spaced learning revision lessons, we hypothesized that intra-lesson 10-min spaces may be beneficial for improving student attainment and engagement in the lessons. The current study, therefore, also investigated the use of a short spacing protocol (10 min), because if this could produce a spacing effect equal to that of a longer strategy, it would be appealing and feasible for schools.

Considering all this evidence and to test the current state of the literature regarding its feasibility (including pupil engagement) for using spaced learning in high school science classroom revision, the present study involved the investigation of three models of spaced learning. One model informed by the cognitive psychology literature featuring longer spaces (24-h spaces between sessions). One model informed by the neuroscience literature with shorter spaces (10-min spaces within sessions), and a combined approach (10 min within session, 24 h between sessions) integrating both sources of evidence.

## Research questions

The overall aim of this project is to draw on the findings from the different literatures on the spacing effect and produce an educational program that is evidence based (with an optimal ISI for revision), but also one that is feasible and engaging for use with students in real world classrooms. Specifically, this study has two main research questions (RQ):

RQ 1. What model of inter study intervals (24h, 10-min, or 24/10) shows the most promise on improving attainment outcomes in GCSE science exam revision classes (i.e., testing a spaced learning theory of intervention)?

RQ 2. Is student engagement with spaced learning revision a significant predictor of pupil science attainment (i.e., testing a spaced learning theory of implementation)?

## Materials and methods

### Study design

The research reported in this manuscript refers to one aspect of a larger study that had three sequenced phases to develop an optimal spaced learning program for revision purposes. The phases were: (1) A design phase; (2) A feasibility pilot phase; and (3) an optimization/comparison phase. However, due to space constraints in this article only the final phase (3. Optimization/comparison) is fully reported. A full report of the first two phases (1. design and 2. feasibility pilot) are provided elsewhere (O'Hare et al., 2017). For the purposes of information and clarity a summary of the design and pilot feasibility phases are provided below.

The first phase (1. design) was a series of program design workshops, which were held between teachers and researchers (cognitive psychologists and neuroscientists) to develop a logic model for high school science revision program using a spaced learning format. This co-design process between researchers and teachers was used to ensure the program had both issues of evidence, feasibility and engagement at the center of its design. The main outcomes of the discussion were that the teachers were already feasibly using 10-min spaces in their revision lessons and the researchers indicated that the cognitive psychology research evidence would suggest that longer spaces, of 24 h or more, have been found to be more effective in memory formation. The discussion culminated in agreement that it was useful to develop several models of a revision program that had combinations of 24 h and 10-min inter study intervals. Three resultant models were produced: one which had 10-min spaces, one with 24 h spaces and one that had 24 h and 10-min combined spaces (see Table 1). The co-design team produced a draft program manual and training materials for the different spaced learning models. The team also designed the content based on the teachers' knowledge of the students' likely level of understanding and the focus of the UK science curriculum.

Phase 1 was followed by a qualitative feasibility pilot study (phase 2) which saw program materials and the three emergent models derived in the first phase (Table 1), piloted in schools to see if they were feasible to deliver in actual classrooms. A control condition using the slides, but no spaces was also piloted for feasibility. The spaced learning models and control condition were piloted in a small number of schools ( $n = 4$ ). Focus groups with  $n = 5$  pupils per school and interviews with  $n = 4$  teachers were used to gain feedback on the feasibility of the different models. Adaptations to training materials and lesson content were made based on this feedback (see O'Hare et al., 2017 for specific changes made between Phase 2 and Phase 3). The outcome measure was also piloted in this

TABLE 1 The three versions of the SMART Spaces program trialed in this study.

	10 min model	24 h model	24/10 model
Day 1	12 min of chemistry 10-min "space" 12 min of chemistry repeated 10-min "space" 12 min of chemistry repeated	12 min of chemistry 12 min of physics 12 min of biology 20 min of "space" at end	12 min of chemistry 10 min of "space" 12 min of physics 10 min of "space" 12 min of biology
Day 2	As day 1 but for physics	As day 1	As day 1
Day 3	As day 1 but for biology	As day 1	As day 1

second phase with two classes per school across the four schools to ensure usability and appropriate timing for the evaluation in Phase 3. Further detail on this measure is outlined in “Measures” below.

The three models in Table 1 have different origins in the literature with the 10-min model emerging from the evidence from the neuroscientific literature (and the practice of teachers involved in this project); the 24 h model incorporating the evidence from the cognitive psychology literature; and the 24/10 model using evidence from neuroscientific and cognitive psychology literature as well as the current teachers’ practice.

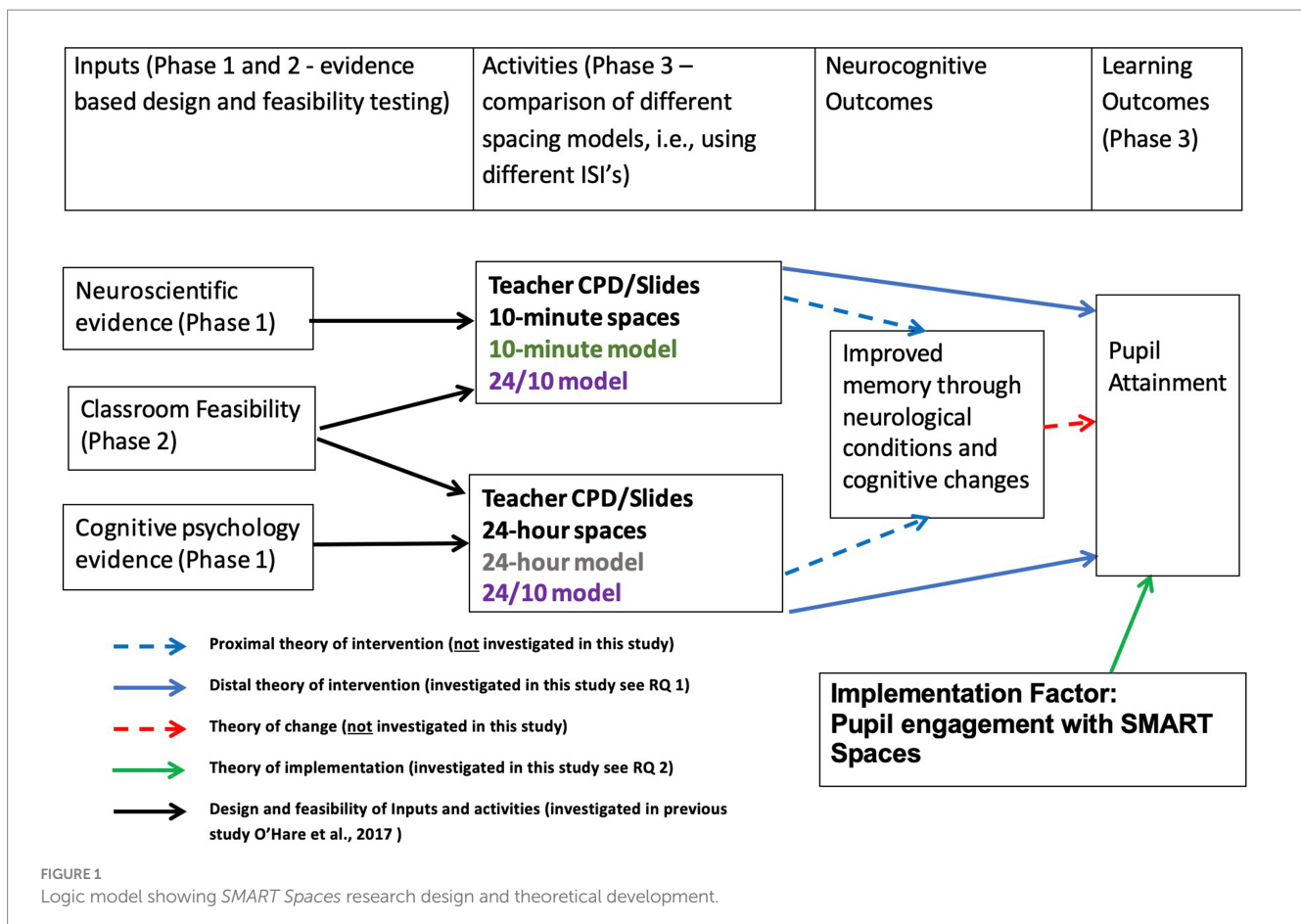
The third phase (and focus of this article) was an optimization pilot randomized controlled trial (RCT). This study is called a “pilot RCT” because the study is not a fully powered effectiveness RCT study. This study was a process to find the optimal ISI model (from phase 2) for use in future classroom applications and investigation through fully powered RCT effectiveness studies. It was never intended to have a fully powered sample size based on a sample size calculation with the required participant numbers to identify effectiveness with a high degree of statistical power. The pilot RCT design was used to give the different ISI models from phase 2 a “fair test” against each other and controls. In addition, it is not a blinded, or double blinded, RCT as it is not possible to hide the method of intervention from trainers, teachers and pupils etc. in educational trials. In fact, it is arguably detrimental to effectiveness if stakeholders are not aware of the program’s theory of intervention.

The three models emerging from phase 2 were compared by being trialed against two control types, namely:

*Control 1* “slides only” was a control group that included the PowerPoint slides but no spaces, i.e., a control of the lesson materials.

*Control 2* was a “no slides or spaces” control that had no materials presented or spaces in their learning, i.e., pupils and teachers received no intervention and carried on with their normal teaching/learning.

All this work from phase 1, 2, and 3 is summarized in the SMART Spaces logic model in Figure 1. This logic model shows how neuroscience and cognitive psychology evidence, along with classroom feasibility, was used to design the three spacing models. It also shows how the optimization study was set up to test each model’s effect on attainment (distal theory of intervention RQ1—see Connolly et al., 2017 Chapter 2 for description of different types of program theory). The logic model also shows that pupil engagement with the program was explored as an indicator of implementation success (theory of implementation RQ2). The study does not elucidate on how the program activities impact upon memory through neurological conditions and cognitive changes (proximal theory of intervention) or how these neurocognitive changes interact with attainment (theory of change). This would require more lab-based or highly controlled conditions featured in much of the previous research. Rather this study focused on the practical questions of optimization of the ISI and program engagement in real world classroom exam revision for improving pupil attainment.



## The SMART Spaces program

This research article explains the experimental optimization of a secondary school science attainment program called *SMART Spaces* which is described in a Template for Intervention Description and Replication (TIDieR—Hoffmann et al., 2014) checklist in Table 2.

The TIDieR checklist details the two main elements to the *SMART Spaces* program, i.e., *SMART Materials* and *SMART CPD* (continuing professional development). The *SMART Materials* comprise of a manual, condensed PowerPoint slides and an activities pack. The manual is a comprehensive guide to the *SMART Spaces* program and is intended to help teachers deliver the program with fidelity (that is, in a manner consistent with the original design) in

any classroom. The manual covers the following elements: background evidence relating to the program's development, the program logic model, the slides for teachers to use during the sessions (chemistry, physics and biology GCSE content), and a step-by-step guide on how to deliver the program. The spacing activities pack is a set of materials that are used in the 10-min "spaces" (distraction activity resources, e.g., juggling balls), and includes a description of how to conduct the activities in various classroom settings.

The *SMART CPD* consists of a half-day CPD course with an experienced teacher in the delivery of the program (usually a GCSE science teacher). *SMART CPD* is a prerequisite for all teachers delivering the program. It includes the presentation of some of the supporting evidence from neuroscience and cognitive psychology, but

TABLE 2 TIDieR checklist for *SMART Spaces* models shared elements.

Brief name	
1	SMART Spaces: Spaced-learning for Memory And Retention Teaching (GCSE Science Revision Version)
Why	
2	Educational program for GCSE students primarily used for examination revision with an aim to improve science attainment for Year 9 and 10 pupils in English schools
What	
3	Materials: SMART Materials, PowerPoint slides, SMART Spaces manual, and SMART Spaces activity pack
4	Procedures: SMART CPD—teachers are trained on delivery of SMART Spaces in a one-day CPD session. SMART Spaces inter-study-interval model (N.B. this is the key aspect being optimized in this study) 3 models were trialed (see Table 1)
Who provided	
5	SMART CPD was provided by trainers experienced in the delivery of SMART Spaces. The same teacher should provide the whole session of SMART Spaces on the three consecutive days
How	
6	Whole-class program that is conducted during three normal science lessons
Where	
7	SMART CPD conducted in out-of-school session, and SMART Spaces lessons are conducted in standard GCSE classroom
When and how much	
8	The program covers GCSE science curriculum content in a high intensity way. The SMART Spaces slides are set out in three 12-min chunks of GCSE chemistry, physics and biology (approximately one third of each course) content to be taught in 1-h lessons, repeated on three consecutive days (see Table 1)
Tailoring	
9	The program logic model was designed using neuroscientific evidence, cognitive psychology evidence, and educational practice literature in both areas. Feasibility was piloted in four schools and optimization was achieved by comparing three different models against two controls in a 12 school RCT (detailed in this manuscript)
Modifications	
10	The optimization study explored the different types of spaces (inter-study intervals) that could be used in delivery of SMART Spaces. It found that there was a clear benefit to using a combination of both 10-min and 24h spaces in the delivery of SMART Spaces content (24/10 SMART Spaces model). Some minor adaptations were made to the inter-study activities (for example, alternative tasks to juggling). No major adaptations are recommended to the emerging model with most promise of 24/10 model
How well	
11	Planned: effective implementation required CPD for teachers in all 12 schools before they took part in the optimization trial and all delivered their assigned version of the program. This CPD was planned to consist of modeling, practice, and feedback on program delivery
12	Actual: the rationale behind this study was to look at the impact of variability in implementation. Among those schools who delivered the same model there was no apparent difference in implementation. The content of the SMART Spaces program was found to have a significant benefit over a no slides/spaces control in the optimization study, with ES $g=0.19$ on total scores on an attainment test using past GCSE questions. Pupil engagement with SMART Spaces was found to be a significant mediator of outcome change. Therefore, it was deemed that SMART Spaces should include the following key elements: 10-min and 24h intervals, SMART CPD, and SMART Resources

TABLE 3 The characteristics of the 12 schools involved in optimization study.

School	Variant type	Urban/rural	Size of school (n)	Free school meals (proxy for disadvantage) pupils (%)	Type of establishment	Ofsted (UK government) quality rating
A	Control 1 (No spaces control)	Urban	1,755	15	Academy Converter	Good
B	10 min model	Rural	1,522	3	Academy Converter	Outstanding
C	24/10 model	Urban	725	15	Community School	Outstanding
D	Control 2 (No slides or spaces control)	Urban	525	21	Academy Converter	Good
E	24/10 model	Urban	1,650	15	Academy Converter	Good
F	24/10 model	Urban	980	21	Academy Converter	Good
G	Control 1 (No spaces control)	Urban	1,244	15	Academy Converter	Good
H	24 h model	Urban	1,600	10	Community School	Good
I	10 min model	Rural	600	16	Academy Converter	Good
J	10 min model	Urban	600	12	University Technical College	Good
K	24 h model	Urban	1,365	35	Academy Sponsor Led	Good
L	Control 2 (No slides or spaces control)	Urban	160	0	Other Independent School	N/A

the major component is modeling how the program is delivered, as well as practice and formative feedback for the teachers on their delivery of *SMART Spaces*. Specifically, the CPD schedule is:

- The scientific background to *SMART Spaces*—the how and the why of why it works (20 min);
- How the sessions are managed, including managing the activities in the “spaces” (15 min);
- A look at the lesson resources provided (15 min);
- An experience of how a *SMART Spaces* session runs (20 min); and
- The opportunity to have a go at delivering a session to the other delegates with constructive formative feedback (20 min per teacher).

## Methods

### Sample, recruitment, and randomization

Recruitment advertisements were shared across England on the funders website, but most schools were recruited through the delivery team's networks in Northern England. There were no selection criteria schools other than they had not previously implemented spaced learning practice in the school. 12 schools agreed to take part and the delivery team responded to interested schools with further information. Each school submitted an expression of interest, and none were excluded from the study as they met the criteria. As explained in study design (page 12) this is not a fully powered RCT sample hence description of the study as a “pilot RCT.” So sample size was based on engagement from eligible schools. All schools were in the Yorkshire and Lancashire area of England and most had a high percentage of pupils in receipt of free school meals (a proxy measure

TABLE 4 Numbers of schools and pupils assigned to each condition.

Condition	School N	Pupil n
10 min model	3	110
24 h model	2	75
24/10 model	3	91
Control 1 (No spaces control)	2	79
Control 2 (No slides or spaces control)	2	53
Total	12	408

for disadvantage). Characteristics of the 12 participant schools are provided in Table 3.

School and pupil numbers for each condition are shown in Table 4. Pupils were all from the same academic year of schooling across all schools (Year 10, i.e., aged 14–15 years).

Randomization was conducted at the school-level. Schools were ordered in terms of numbers of participants and then divided into two groups based on participant numbers (Group A = six schools with largest participant numbers, Group B = six schools with smallest participant numbers). Random numbers were generated for each group to allocate them to one of the five conditions. The remaining schools, one in each group, were allocated to the 10-min and 24-h variants, respectively, (to ensure some participants in these two variants in case of school withdrawal). Four schools were pre-tested after randomization due to practical time constraints. Randomization took place on 23 February 2016; four schools were pre-tested in a window 2 weeks after that, up to 8 March 2016. One school (School L, a small independent school) assigned to the 24-h space variant did not wish to take part in the training and delivery of the program as assigned, but agreed to the pre- and post-test (with 11 of the 14 pupils providing complete data).

TABLE 5 Pre-test, post-test for all delivery models and controls.

Variant		Mean	N	Std. deviation	Std. error mean
10 min model	Pre-test - Total score (max = 57)	27.93	100	9.16	0.92
	Post-test - Total score (max = 57)	30.82	100	9.82	0.98
24 h model	Pre-test - Total score (max = 57)	25.77	69	12.94	1.56
	Post-test - Total score (max = 57)	27.30	69	14.64	1.76
24/10 mode	Pre-test - Total score (max = 57)	23.79	76	10.87	1.25
	Post-test - Total score (max = 57)	27.55	76	12.22	1.40
Control 1 No spaces	Pre-test - Total score (max = 57)	21.37	68	8.70	1.05
	Post-test - Total score (max = 57)	23.99	68	8.89	1.08
Control 2 No slides or spaces	Pre-test - Total score (max = 57)	25.00	50	7.20	1.02
	Post-test - Total score (max = 57)	26.82	50	9.80	1.39

Therefore, it was reassigned to the “no spaces/materials” control group. This reassignment violates full RCT intention-to-treat characteristics, but the study is still an RCT. However, as feasibility and ISI optimization were the foci of the study showing evidence of program promise rather than actual efficacy of the program, it was deemed appropriate by the evaluation team to include this school in the analysis.

All AQA GCSE<sup>1</sup> science pupils in the schools were eligible for the program on the condition that their class teachers had received the *SMART Spaces* CPD. Classes of pupils were chosen within each participating school by the project contact for each school. A teacher who returned an expression of interest may have volunteered their own teacher time, and may have asked other teachers to also participate in the study. Schools did not include all GCSE science pupils—one to two classes were chosen in each school by the participating teachers (there was a total of 408 pupils across all schools). All research was conducted according to (Queen's University Belfast) School of Education ethical guidelines. Ethical consent was obtained from the Ethics Committee before data collection was conducted. Informed consent was sought at the pupil level through opt-out consent forms (sent home to parents and verbally explained to pupils at testing) for informed participation in the program, and completion of pre-tests and post-tests. The data collected was coded and entered onto a database, anonymized, and held securely on a password-protected computer.

## Measures

The main outcome measure was a bespoke secondary school science test, comprising past-paper questions from the AQA GCSE curriculum. The questions were selected by the research team from a range of past papers. The teacher delivery team were blind to the content of the outcome measure; so as not to influence the content of the CPD sessions or encourage adaptation of slides to include additional emphasis on the exam questions used. A reliability analysis

of the test showed a Cronbach's alpha = 0.88. This test had two sections: Section A—short answers and multiple choice, and Section B—long answers. There were 39 marks available for Section A and 18 marks for Section B—a total maximum score of 57. The short answer and multiple-choice section (Section A) required participants to give answers ranging from one word to two or three lines; the long answer section (Section B) required considerably more detail per answer, requiring five to six key points of information. The test had a time limit of 45 min.

Data from teacher focus groups provided during phase 2 of the study (not reported here—see O'Hare et al., 2017 for more details) reported that engagement of pupils in the program was an important implementation factor. This data was used to design an implementation questionnaire which was administered to pupils post-implementation. There were 13 items in the pupil engagement scale and reliability of this measure was very good (Cronbach's alpha = 0.91).

Mean retention interval (i.e., lag between pre-test and post-test) was 18 days (SD = 12). The retention interval varied across schools due to availability of the school for a testing visit from the research team. We controlled for retention interval using a regression model (including pre-test score as a predictor) and retention interval was not predictive for post-test scores. This natural variation of retention interval, around 3 weeks is representative of when schools would use a revision intervention.

## Analysis

### Comparison of model effects on improving attainment (research question 1)

To investigate the relative effects of the three different spacing models, independent *t*-tests were used to compare pupil's attainment gain scores (difference between their pre and post-test) for each model with the two controls. Attainment gain scores were used to control for baseline score (see Table 5 for mean pre, post and attainment gain scores for conditions). For example, pupils' gain scores in the 10-min group were compared (using an independent *t*-test) with pupil gain scores in Control group 1 (“slides only” control). In total there are six comparisons, i.e.: 10-min model with Control 1 and Control 2; 24-h model with

<sup>1</sup> AQA is a U.K. exam board, and GCSEs (“General Certificates of Secondary Education”) are national subject-specific awards typically taught and conducted in the U.K. in Years 10 and 11 (age 15–16years).



TABLE 6 Pre-post paired *t*-tests for all groups with effect size of gain on total score for the three models compared to the two control groups.

Spaced learning models and controls (N)	Mean difference pre-post	S.D.	S.E.	95% confidence interval of the difference		<i>t</i>	df	Paired <i>t</i> sig. (2-tailed)	Effect compared to control type	Effect size = <i>d</i> (95% confidence intervals) <sup>a</sup>
				Lower	Upper					
10 min model (100)	2.89	5.27	0.57	1.84	3.94	5.49	99	<0.01	No spaces	0.03 (−0.14, 0.20)
									No slides or spaces	0.11 (−0.07, 0.3)
24 h model (69)	1.54	5.40	0.65	0.24	2.83	2.36	68	0.02	No spaces	−0.09 (−0.25, 0.07)
									No slides or spaces	−0.02 (−0.20, 0.15)
24/10 model (76)	3.76	5.01	0.57	2.62	4.91	6.55	75	<0.01	No spaces	0.11 (−0.05, 0.27)
									No slides or spaces	0.19* (0.01, 0.36)
Control 1: No spaces (68)	2.62	4.92	0.60	1.43	3.81	4.39	67	<0.01		
Control 2: No slides or spaces (50)	1.82	5.19	0.73	0.35	3.29	2.48	49	0.02		

\*Significant < 0.05. <sup>a</sup>Standardized Mean Difference (*d*) calculated from mean gain scores, pre and post SDs, and paired *t*-tests available at Campbell Collaboration calculator: <https://www.campbellcollaboration.org/research-resources/effect-size-calculator.html>.

Control 1 and Control 2; and 24/10 Model with Control 1 and Control 2. From the results of these independent *t*-tests the effect sizes (ES = Cohen's *D*) were produced using standardized mean difference (*d*) calculated from mean gain scores, pre and post SDs, and paired *t*-tests (effect size calculator available at Campbell Collaboration<sup>2</sup>).

Finally, it is important to consider the educational significance of effect sizes in education trials. Kraft (2020) analyzed the distribution of 1942 effect sizes from 747 education RCTs and determined classifications of effect sizes: less than 0.05 = small effect, 0.05 to less than 0.20 = medium effect, and 0.20 or greater = large effect. Bloom et al. (2008) found that by age 10, children's educational achievement progresses by an effect size of 0.4 per year, and as such, effect sizes in education trials have a threshold lower than has traditionally been interpreted.

There is debate over when correction for multiple comparisons should be applied to correct for possible inflated risk of Type I errors. We argue that the present study meets the criteria for not requiring correction for multiple comparisons, as all analyses were pre-planned (Armstrong, 2014). The design of the study, comparing multiple variants of spaced learning, is indicative of this pre-planning—the multiple comparisons are of these variants specifically. Correcting for multiple comparisons unnecessarily may itself present an inflated risk of Type II error (Gelman et al., 2012). However, in appreciation of the uncertainty of this debate, we have also presented an alternative analysis, using a one-way ANOVA of pre-test to post-test gain scores for a between-groups factor of variant, thus providing a single significance test for the effect of variant on gain scores. We followed this with Bonferroni-corrected *post-hoc* tests, the most conservative-correction for multiple comparisons. Finally, we also present Tukey-corrected *post-hoc* tests, still correcting for multiple comparisons, but with less risk of Type II error.

## Exploration of pupil engagement as an implementation factor (research question 2)

Two hundred and twenty-four pupils received one of the three models of the program (i.e., were not in either control group). These pupils completed pre- and post-tests and the post-test engagement questionnaire. A sub-group regression analysis was conducted using only data for these pupils to investigate the relationship between their attainment gains and engagement.

## Results

### Comparison of model effects on improving attainment (research question 1)

The independent *t*-tests for the 24/10 model showed a consistent pattern of positive effects when compared to both controls (with a positive effect indicating improved performance of the intervention group over the control group) (Table 6). One of these effects was significant between the total gain score of the 24/10 model and the total gain score of “no slides or spaces” Control 2 (ES = 0.19). There was a more modest pattern of positive effects of the 10-min model compared to controls, but with no significant effect. The 24 h model produced negative effects in comparison to the two controls. Therefore, it can be seen that 24/10 model shows the most consistent evidence of promise against controls at this stage of program development. This effect of 0.19 is at the upper limit of Kraft's (2020) medium category of effect sizes: effects of 0.05 to less than 0.20 = medium, 0.2 or greater are “large.” It should also be noted that all the spaced learning models performed better against the “no slides no spaces” control rather than the “no spaces” control. Thus suggesting an intrinsic benefit of the slides in themselves.

As discussed in the methodology, we present an alternative analysis in appreciation of the debate over multiple comparisons and error risks: a one-way ANOVA for the effect of variant on pre-test to post-test gain score for Total score. This gave a result of  $F(4,358) = 2.058$ ,  $p = 0.086$ , suggesting that the overall effect of variant on gain score is non-significant.

<sup>2</sup> <https://www.campbellcollaboration.org/research-resources/effect-size-calculator.html>

We followed this with Bonferroni-corrected *post-hoc* comparisons of each pair of variants (Table 7) which showed that no comparisons of any variant with another was statistically significant when this correction was applied.

Finally, Table 8 shows Tukey HSD *post-hoc* tests for each pair of variants effect on gain score, which shows the only comparison approaching significance, is for 10 min 24 h variant having a higher mean gain score than 24 h variant ( $p=0.07$ ).

## Exploration of pupil engagement as an implementation factor (research question 2)

The pupil engagement score was a significant implementation predictor, with higher engagement scores predicting more positive outcome change (the adjusted R Square for the model was 0.81 showing the high degree of the variance in post-test score being predicted by pre-test and engagement score). Looking at the standardized co-efficients it can be seen that the vast majority of the variance in the post-test score is predicted by the pre-test score ( $b=0.40$ ) compared to the engagement score ( $b=0.06$ ) but engagement (controlling for pre-test score) is still a significant predictor of performance and an implementation variable that teachers can influence, thus worthy of note (Table 9).

## Discussion

Previous literature on the spacing effect has shown the benefits of both short spaces (around 10 min—Kelley and Whatson, 2013) and medium-term spaces (24 h plus—Cepeda et al., 2006) on the retention of information. Furthermore, some educational practice literature shows the benefits of both these kinds of spaces in real-world educational settings (Dunlosky et al., 2013). The emerging picture from this research would be consistent with that literature. Also, when comparing different spacing models (as described in the “The SMART Spaces program” section above) this research suggests that there is promise to combining both short and medium-term spaces for improving attainment outcomes. This has resulted in the 10-min and 24-h spacing pattern (24/10 model) underpinning SMART Spaces.

Regarding SMART Spaces theoretical development, it is useful to reflect on the logic model (Figure 1). The previous work (O'Hare et al., 2017) showed how a mix of research evidence and feasibility study could generate models of spaced learning that can be applied during classroom revision and tested for their impact on attainment. Generally, the 24/10 model fits within the constraints of most school timetables as it enables delivery that can be completed within an hour. Programs taking more than this time require greater re-organization within a school and so are often beyond the means of an ordinary classroom teacher. Also, student engagement was found to be high during the lessons, which is important as engagement was found to be a significant implementation predictor of attainment outcomes.

TABLE 7 Bonferroni-corrected *post-hoc* tests for comparison of variant effect on gain pre-test to post-test gain scores for total score.

Variant	Variant	Mean difference	Std. error	Sig.	95% confidence interval	
					Lower bound	Upper bound
10Min	24Hour	1.35	0.81	0.95	-0.93	3.64
	NoSpaces	0.27	0.81	1.00	-2.02	2.57
	10Min_24Hour	-0.87	0.79	1.00	-3.09	1.35
	Control	1.07	0.89	1.00	-1.46	3.60
24Hour	10Min	-1.35	0.81	0.95	-3.64	0.93
	NoSpaces	-1.08	0.88	1.00	-3.57	1.41
	10Min_24Hour	-2.23	0.86	0.10	-4.65	0.20
	Control	-0.28	0.96	1.00	-2.99	2.43
NoSpaces	10Min	-0.27	0.81	1.00	-2.57	2.02
	24Hour	1.08	0.88	1.00	-1.41	3.57
	10Min_24Hour	-1.15	0.86	1.00	-3.58	1.29
	Control	0.80	0.96	1.00	-1.92	3.52
10Min_24Hour	10Min	0.87	0.79	1.00	-1.35	3.09
	24Hour	2.23	0.86	0.10	-0.20	4.65
	NoSpaces	1.15	0.86	1.00	-1.29	3.58
	Control	1.94	0.94	0.40	-0.71	4.60
Control	10Min	-1.07	0.89	1.00	-3.60	1.46
	24Hour	0.28	0.96	1.00	-2.43	2.99
	NoSpaces	-0.80	0.96	1.00	-3.52	1.92
	10Min_24Hour	-1.94	0.94	0.40	-4.60	0.71

Based on observed means.  
The error term is Mean Square (Error) = 26.676.

TABLE 8 Tukey HSD *post-hoc* tests for comparison of variant effect on gain pre-test to post-test gain scores for total score.

Variant	Variant	Mean difference	Std. error	Sig.	95% confidence interval	
					Lower bound	Upper bound
10Min	24Hour	1.35	0.81	0.45	-0.86	3.57
	NoSpaces	0.27	0.81	1.00	-1.95	2.50
	10Min_24Hour	-0.87	0.79	0.80	-3.03	1.28
	Control	1.07	0.89	0.75	-1.38	3.52
24Hour	10Min	-1.35	0.81	0.45	-3.57	0.86
	NoSpaces	-1.08	0.88	0.74	-3.50	1.34
	10Min_24Hour	-2.23	0.86	0.07	-4.58	0.13
	Control	-0.28	0.96	1.00	-2.91	2.35
NoSpaces	10Min	-0.27	0.81	1.00	-2.50	1.95
	24Hour	1.08	0.88	0.74	-1.34	3.50
	10Min_24Hour	-1.15	0.86	0.67	-3.51	1.22
	Control	0.80	0.96	0.92	-1.84	3.44
10Min_24Hour	10Min	0.87	0.79	0.80	-1.28	3.03
	24Hour	2.23	0.86	0.07	-0.13	4.58
	NoSpaces	1.15	0.86	0.67	-1.22	3.51
	Control	1.94	0.94	0.24	-0.64	4.52
Control	10Min	-1.07	0.89	0.75	-3.52	1.38
	24Hour	0.28	0.96	1.00	-2.35	2.91
	NoSpaces	-0.80	0.96	0.92	-3.44	1.84
	10Min_24Hour	-1.94	0.94	0.24	-4.52	0.64

TABLE 9 Regression of independent variables pre-test and engagement onto post-test outcome scores.

Independent variables	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. error	Beta		
(Constant)	-0.97	2.04		-0.47	0.64
Pre-test total score (max = 57)	0.99	0.03	0.40	30.54	<0.01
Engagement score	0.11	0.05	0.06	2.13	0.03

Dependent variable: post-test total score ( $N = 224$ )<sup>a</sup> across all model types (max score = 57).

<sup>a</sup>Note this sample size is smaller than the feasibility study overall sample size as it is a sub-group analysis of the three spaced learning conditions.

An observed pattern in the current study was that all the models performed better against the “no slides and no spaces” control group rather than the “slides-only” group (see Table 6). This suggests that there is some intrinsic benefit in the way the content is presented. Therefore, it is important that the slides are of high quality and updated regularly based on the current curriculum and the key words that students must use in the exam are clearly elucidated, repeated in context by the teacher and then recalled and practiced by the student. There was less difference between the three spacing models and the “no spaces control” and the “no slide no spaces control,” and so we must acknowledge that the study found evidence of the effect of the program as a whole (i.e., CPD, slides, spaces etc.) and not simply evidence for the benefits of a particular spacing strategy.

Beyond these points, the underpinning theory of intervention is that the *SMART Spaces* 24/10 model proximally improves memory through neurological conditions and cognitive changes which have a

distal effect on pupil attainment. This study only investigated the distal effects of the 24/10 model on attainment (compared against other models). The neurocognitive changes are only hypothesized at this point as a proximal theory of intervention. New neuroscientific and cognitive study would be required to investigate this proximal theory of intervention as explored in the “Inter-study Intervals and Attainment” section above. For example, sleep (Bell et al., 2014) and regeneration or proteins such as CREB (Fields, 2009) are potentially fruitful areas of future investigation.

Taking a wider view, there may be criticism of the 24/10 model in the educational community as this approach focuses on the acquisition of facts and key points in a defined set of science topics, rather than focussing on deeper learning and the application of science knowledge in practical contexts. We must acknowledge some key counterpoints to this criticism. Firstly, for students who have not achieved successful retention of key facts yet, doing so during

revision could be extremely beneficial. Secondly, covering the key facts in a time-efficient manner may allow students to then move on to practical applications at an earlier stage and spend more time on these other facets of science education. Finally, the program may be criticized for focussing its spacing strategy on mid-length retention rather than long term (as would be encouraged if we used a spacing protocol of weeks or months). But being pragmatic about revision, the goal is exam performance, and a 24-h space may therefore be most appropriate when the examination is in the near future. This is supported in the earlier literature (Cepeda et al., 2006) and as interpreted above in the “Inter-study Intervals and Attainment” section. Furthermore, some students, particularly those in disadvantaged circumstances, may only revise in a school context and therefore it is important that the revision conducted within school is as effective as possible.

A methodological point is that this research project is an example of the benefits of conducting pilot work and small-scale trial studies as well as using theory and evidence to inform the design of educational programs, rather than prematurely moving to large RCT type studies of interventions, or going to large scale program implementation prematurely. Furthermore, it demonstrates the benefits of conducting this pilot work in a research and practice partnership (i.e., where teachers and researchers work together to co-design or co-construct an educational program) for easier integration of evidence and ensuring feasibility and pupil engagement in real world classroom settings.

Regarding application of these findings, although we have designed and tested the 24/10 *SMART Spaces* model in a secondary school science classroom there may be applications of it for enhancing attainment in other environments. Arguably the model could be easily applied, for examination and revision purposes, to other school subjects (languages, mathematics etc.) at other levels of education (e.g., elementary and middle school). However, there are also a wide range of contexts outside the school classroom that it could help improve performance (e.g., healthcare and industry settings). Training for many jobs requires role specific content to be learned quickly and yet well remembered. The 24/10 *SMART Spaces* model could potentially add efficiency and cost effectiveness in these situations. It is important to consider the measured effects in the context of changes in examination performance. The effect of the 24/10 model in the above analysis equates to 4–5% of an increase in test score. Considering UK GCSE examinations are graded on a 9 point scale, this could substantially shift a student toward a higher grade boundary, especially if this strategy was applied across all examination content and gains could be realized across the multiple papers a student must sit for science GCSE qualifications. Furthermore, an intervention like *SMART Spaces*, which has a comparatively low cost and low commitment for teacher's, is arguably a productive use of teacher time if the low and medium effect sizes (based on Kraft, 2020) found in this research are replicated in future research.

These potential gains, however, must be considered alongside the caveat that this is an early stage pilot evaluation, and not a fully powered RCT. Some limitations must, therefore, be considered. First, although the *t*-tests (Table 6) showed a significant effect of the 24/10 model these significant effects were not apparent when the corrections of multiple comparisons was included. We have argued that this correction is not appropriate in this case as the comparisons

were pre-determined before analysis. However, on balance the lack of significant effects in the *post hoc* tests would indicate any potential effects of spaced learning in this application are fairly weak. Another limitation is the sample size is relatively small for an RCT, and not adequately powered to confidently detect significance in the expected effects. The number of schools and pupils per condition was also small and varied substantially. This variation in numbers per condition, or an anomalous school in terms of implementation quality, could have had an undue influence on the effect size of the model being delivered. Although there is not a particularly strong weight of evidence to select the 24/10 model over the others, the nature of a pilot study is to investigate what is most promising, and the quantitative effects, in combination with the qualitative feedback and student engagement data, suggest that this is the model most likely to succeed from our piloted variations. Finally, data were not gathered on what constituted control group activity, i.e., what “business-as-usual” involved. This introduces a limitation when comparing the spaced learning conditions with this control group, as it is possible the control group used other effective revision strategies or their own independent attempts at spaced learning. The moderately low level of pre-test to post-test changes for the no slides or spacing control group does not suggest that there was usage of any efficient revision strategy, but this must be considered in future work and rigorous examination of contamination or other relevant control group activity should be analyzed. Regardless, if the control groups were using some revision strategies then this would have dampened the effects found for the 24/10 model rather than inflating them.

Future research would require a larger sample size and comparison of fewer variations of the program (ideally control and intervention). Future research would also benefit from being in a real-world context, i.e., as revision program before a national standardized exam such as GCSE in the UK. This real-world test would also need to consider the presence about current revision practices used in schools, e.g., past papers, and whether 24/10 *SMART Spaces* model adds value or alters these practices. In fact, the study presented here is succeeded by a large-scale efficacy trial of the *SMART Spaces* program (see Hodgen et al., 2018 for a research protocol) and will explore all these issues in more detail and at a greater scale. Finally, as previously mentioned more work is needed to understand the proximal theory of intervention in terms of the cognitive and neurological changes that occur as a result of the spacing effect.

## Conclusion

The main aim of this study was to compare several models of spaced learning for their effects on educational attainment when used during examination revision. The most promising model used 24-h spaces between repetitions of science material, with 10-min breaks within each repetition (the *SMART Spaces* 24/10 model), which was consistent with effectiveness studies in the neuroscience and cognitive psychology literature on the spacing effect. The present findings demonstrate that (1) the spacing effect can be utilized feasibly in the classroom for revision purposes, and (2) the *SMART Spaces* program shows promise as a specific way to use the spaced learning through revision to improve attainment.

The key theoretical intervention mechanism at the heart of this program, i.e., the spacing effect, has had a century of evidence behind

it, yet explicit, evidence informed and classroom-based implementation of it remains scarce. Ultimately, high quality revision strategies are obviously of value in schools, and this paper provides evidence that the 24/10 *SMART Spaces* model should be a strategy to consider for future classroom use and research.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by School of Social Sciences Education and Social Work at Queen's University Belfast. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because the ethics committee approved that informed consent could be sought at the pupil level through opt-out consent forms sent home to parents/guardians and returned if they wished their child's pre and post test data not to be included. Also, consent was explained verbally to pupils at testing for informed participation in the program, and completion of pre-tests and post-tests. In addition, all participating schools provided written informed consent.

## References

- Abott, E. E. (1909). On the analysis of the factor of recall in the learning process. *Psychol. Rev. Monogr. Suppl.* 11, 159–177. doi: 10.1037/h0093018
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic Physiol. Opt.* 34, 502–508. doi: 10.1111/opo.12131
- Barzagar Nazari, K., and Ebersbach, M. (2019). Distributing mathematical practice of third and seventh graders: applicability of the spacing effect in the classroom. *Appl. Cogn. Psychol.* 33, 288–298. doi: 10.1002/acp.3485
- Bell, M. C., Kawadri, N., Simone, P. M., and Wiseheart, M. (2014). Long-term memory, sleep, and the spacing effect. *Memory* 22, 276–283. doi: 10.1080/09658211.2013.778294
- Bird, S. (2011). Effects of distributed practice on the acquisition of second language English syntax. *Appl. Psycholinguist.* 32, 435–452. doi: 10.1017/S0142716410000470
- Bloom, H. S., Hill, C. J., Black, A. R., and Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *J. Res. Educ. Effect.* 1, 289–328. doi: 10.1080/19345740802400072
- Carpenter, S. K., Pashler, H., and Cepeda, N. J. (2009). Using tests to enhance 8th grade students retention of U.S. history facts. *Appl. Cogn. Psychol.* 23, 760–771. doi: 10.1002/acp.1507
- 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. *Psychol. Bull.* 132, 354–380. doi: 10.1037/0033-2909.132.3.354
- Chen, O., Castro-Alonso, J. C., Paas, F., and Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: evidence from the spacing effect. *Educ. Psychol. Rev.* 30, 483–501. doi: 10.1007/s10648-017-9426-2
- Connolly, P., Biggart, A., Miller, S., O'Hare, L., and Thurston, A. (2017). *Using randomised controlled trials in education*. United States: SAGE.
- Donoghue, G. M., and Horvath, J. C. (2016). Translating neuroscience, psychology and education: an abstracted conceptual framework for the learning sciences. *Cogent Educ.* 3:1267422. doi: 10.1080/2331186X.2016.1267422
- 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. *J. Appl. Psychol.* 84, 795–805. doi: 10.1037/0021-9010.84.5.795
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., and Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and Educational Psychology. *Psychol. Sci. Public Interest* 14, 4–58. doi: 10.1177/1529100612453266
- Ebbinghaus, H. (1885). *Über das Gedächtnis: untersuchungen zur experimentellen psychologie*. Duncker and Humblot: Leipzig.
- Fields, R. D. (2009). Making memories stick. *Sci. Am.* 120:5. doi: 10.1126/scisignal.2009ec285
- Firth, J. (2021). Boosting learning by changing the order and timing of classroom tasks: implications for professional practice. *J. Educ. Teach.* 47, 32–46. doi: 10.1080/02607476.2020.1829965
- Friedman, S., and Laurison, D. (2020). *The class ceiling: Why it pays to be privileged*. Bristol, United Kingdom: Bristol University Press.
- Gelman, A., Hill, J., and Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *J. Res. Educ. Effect.* 5, 189–211. doi: 10.1080/19345747.2011.618213
- Gobet, F. (2005). Chunking models of expertise: implications for education. *Appl. Cogn. Psychol.* 19, 183–204. doi: 10.1002/acp.1110
- Gurung, R. A., and Burns, K. (2019). Putting evidence-based claims to the test: a multi-site classroom study of retrieval practice and spaced practice. *Appl. Cogn. Psychol.* 33, 732–743. doi: 10.1002/acp.3507
- Hodgen, J., Anders, J., Bretscher, N., and Hardman, M. (2018). Trial Evaluation Protocol: SMART Spaces (Spaced Learning Revision Programme). *Education Endowment Foundation*. London, UK.
- Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., et al. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 348:g1687. doi: 10.1136/bmj.g1687
- Horvath, J. C., and Donoghue, G. M. (2016). A bridge too far—revisited: reframing Bruer's neuroeducation argument for modern science of learning practitioners. *Front. Psychol.* 7:377. doi: 10.3389/fpsyg.2016.00377
- Janiszewski, C., Noel, H., and Sawyer, A. G. (2003). A Meta-analysis of the spacing effect in verbal learning: implications for research on advertising repetition and consumer memory. *J. Consum. Res.* 30, 138–149. doi: 10.1086/374692

## Author contributions

LO'H: principal investigator, project design, analysis, and write up. PS: project manager, project design, analysis, and write up. AG: practice lead, project design, analysis, and write up. CM, AT, and AB: project design, analysis, and write up. All authors contributed to the article and approved the submitted version.

## Funding

This research was funded by the Education Endowment Foundation.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Kapler, I. V., Weston, T., and Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: long-term benefits for factual and higher-level learning. *Learn. Instr.* 36, 38–45. doi: 10.1016/j.learninstruc.2014.11.001
- Kelley, P., and Watson, T. (2013). Making long-term memories in minutes: a spaced learning pattern from memory research in education. *Front. Hum. Neurosci.* 7:589. doi: 10.3389/fnhum.2013.00589
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educ. Res.* 49, 241–253. doi: 10.3102/0013189X20912798
- Kramar, E. A., Babayan, A. H., Gavin, C. F., Cox, C. D., Jafari, M., Gall, C. M., et al. (2012). Synaptic evidence for the efficacy of spaced learning. *Proc. Natl. Acad. Sci.* 109, 5121–5126. doi: 10.1073/pnas.1120700109
- Mauelshagen, J., Sherff, C. M., and Carew, T. J. (1998). Differential induction of long-term synaptic facilitation by spaced and massed applications of serotonin at sensory neuron synapses of *Aplysia californica*. *Learn. Mem.* 5, 246–256. doi: 10.1101/lm.5.3.246
- Miles, S. W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System* 42, 412–428. doi: 10.1016/j.system.2014.01.014
- Mollison, M. V., and Curran, T. (2014). "Investigating the Spacing Effect Using EEG." In *Temporal Dynamics of Learning Center Annual Meeting*. University of San Diego, San Diego, CA, USA.
- Moss, V. D. (1995). The efficacy of massed versus distributed practice as a function of desired learning outcomes and grade level of the student. *Dissertation Abstracts International: Section B: The Sciences and Engineering* 56:5204. Available at: <http://ezproxy.lib.utexas.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1996-95005-375&site=ehost-live>
- Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Prims, J., Melton, Z. J., et al. (2017). The state of social and personality science: Rotten to the core, not so bad, getting better, or getting worse? *Pers. Soc. Psychol.* 113, 34–58. doi: 10.1037/pspa0000084
- Namaziandost, E., Mohammed Sawalmeh, M. H., and Soltanabadi, M. I. (2020). The effects of spaced versus massed distribution instruction on EFL learners' vocabulary recall and retention. *Cogent Educ.* 7:1. doi: 10.1080/2331186X.2020.1792261
- O'Hare, L. (2014). Did Children's perceptions of an after-school social learning program predict change in their behavior? *Procedia Soc. Behav. Sci.* 116, 3786–3792. doi: 10.1016/j.sbspro.2014.01.842
- 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*. United Kingdom: Education Endowment Foundation.
- O'Hare, L., Stark, P., Orr, K., Biggart, A., and Bonnell, C. (2018). *Positive action. Evaluation report and executive summary*. United Kingdom: Education Endowment Foundation.
- Putwain, D. W. (2008). Supporting assessment stress in key stage 4 students. *Educ. Stud.* 34, 83–95. doi: 10.1080/03055690701811081
- Rawson, K. A., and Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: how much is enough? *J. Exp. Psychol. Gen.* 140, 283–302. doi: 10.1037/a0023956
- Rønning, M. (2011). Who benefits from homework assignments? *Econ. Educ. Rev.* 30, 55–64. doi: 10.1016/j.econedurev.2010.07.001
- Sobel, H. S., Cepeda, N. J., and Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Appl. Cogn. Psychol.* 25, 763–767. doi: 10.1002/acp.1747
- Timmer, M. C., Steendijk, P., Arend, S. M., and Versteeg, M. (2020). Making a lecture stick: the effect of spaced instruction on knowledge retention in medical education. *Med. Sci. Edu.* 30, 1211–1219.
- Van Strien, J. W., Verkoeijen, P. P. J. L., Van der Meer, N., and Franken, I. H. A. (2007). Electrophysiological correlates of word repetition spacing: ERP and induced band power old/new effects with massed and spaced repetitions. *Int. J. Psychophysiol.* 66, 205–214. doi: 10.1016/j.ijpsycho.2007.07.003
- Wiseheart, M., Küpper-Tetzel, C. E., Weston, T., Kim, A. S., Kapler, I. V., and Foot-Seymour, V. (2019). *Enhancing the quality of student learning using distributed practice*.
- Xue, G., Mei, L., Chen, C., Lu, Z. L., Poldrack, R., and Dong, Q. (2011). Spaced learning enhances subsequent recognition memory by reducing neural repetition suppression. *J. Cogn. Neurosci.* 23, 1624–1633. doi: 10.1162/jocn.2010.21532
- Zhang, Y., Liu, R. Y., Heberton, G. A., Smolen, P., Baxter, D. A., and Cleary, L. J. (2012). Computational design of enhanced learning protocols. *Nat. Neurosci.* 15, 294–297.
- Zwaan, R. A., Pecher, D., Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., et al. (2017). Participant Nonnaïveté and the reproducibility of cognitive psychology. *Psychon. Bull. Rev.* 25, 1968–1972. doi: 10.3758/s13423-017-1348-y