Education Endowment Foundation

Spaced Learning: The Design, Feasibility and Optimisation of SMART Spaces

Evaluation report and executive summary

March 2017

Independent evaluators:

Dr Liam O'Hare, Dr Patrick Stark, Professor Carol McGuinness, Dr Andy Biggart, Professor Allen Thurston

Queen's University Belfast

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus – Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

This project was funded as part of the Education and Neuroscience scheme, which was jointly funded by the Wellcome Trust and Education Endowment Foundation and launched in January 2014. The aim of the scheme was to provide funding for collaborative projects between educators and neuroscientists to develop evidence-based interventions for use in the classroom, or to rigorously test existing tools and practices.

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

Danielle Mason
Head of Research and Publications
Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP
t: 020 7802 1679
e: danielle.mason@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

The project was independently evaluated by a team from Queens University Belfast: Dr Liam O'Hare, Dr Patrick Stark, Professor Carol McGuinness, Dr Andy Biggart, Professor Allen Thurston.

The lead evaluator was Dr Liam O'Hare.

Contact details:

Dr Liam O'Hare
Queens University Belfast,
University Road,
Belfast,
BT7 1NN,
Northern Ireland,
U.K.
Tel: +44 (0)28 9097 5973
Email: L.OHare@qub.ac.uk

Centre for Evidence and Social Innovation at Queen's University Belfast
http://www.qub.ac.uk/research-centres/cesi/

Contents

Executive Summary

The project

This report describes the development and pilot evaluation of SMART Spaces. This programme aims to boost GCSE science outcomes by applying the principle that information is more easily learnt when it is repeated multiple times, with time passing between the repetitions. This approach is known as 'spaced learning' and is contrasted with a 'massed learning' approach, where content is learnt all at once with no spacing. The development of the programme was led by a team from the Hallam Teaching School Alliance (HTSA). SMART Spaces prepares Year 9 and 10 students for GCSE examinations at the end of Year 10. Teachers were trained to deliver three lessons focused on chemistry, physics, and biology curriculum content, which were repeated over three consecutive days. Pupils did an unrelated physical activity in the spaces between intensive repetitions of science content. Teachers received one day of training and were provided with PowerPoint slides to deliver during the lessons.

The Centre for Evidence and Social Innovation (CESI) at Queen's University Belfast (QUB) worked with HTSA to develop SMART Spaces, test its feasibility, and test three different approaches to arranging the spaced learning across the three days (see Table 1).

This project was jointly funded by the EEF and Wellcome Trust as part of the Education and Neuroscience partnership.

Key conclusions	
1.	The spaced learning principle is based on relatively strong evidence and this pilot suggests that **SMART Spaces** has evidence of promise.
2.	This pilot study demonstrated that **SMART Spaces** can feasibly be delivered in English schools. Both teachers and pupils appeared to enjoy and engage with the programme.
3.	Teachers generally reported that they delivered **SMART Spaces** lessons as prescribed and did not make major alterations.
4.	The small randomised controlled trial (RCT) provided some preliminary evidence that the most promising approach to spaced learning combines the use of both ten-minute and 24-hour spaces between curriculum content. However, this was a small study and a larger trial is needed to better understand the impact of the programme.
5.	**SMART Spaces** is ready for a larger RCT to evaluate its impact on GCSE attainment.

What are the findings?

The principle of spaced learning is supported by evidence from two scientific fields, neuroscience and cognitive psychology. The neuroscience literature supports the use of shorter spaces between learning (of around ten minutes), and the cognitive psychology literature supports longer spaces (of around 24 hours). This study used a small randomised controlled trial to investigate the efficacy of three different versions of the SMART Spaces programme (see Table 1). This suggested that Version 3—which combined 24-hour and ten-minute spacing—appeared to be the most promising variant. This version was also supported by *both* the neuroscientific and cognitive psychology literature, and will be used in future implementations of SMART Spaces. However, this study was only intended to provide preliminary evidence and is smaller than EEF efficacy trials. A larger trial is required before any firm conclusions can be drawn about the efficacy of SMART Spaces.

Table 1: the three versions of the SMART Spaces programme

	Version 1 (10-minute spaces)	Version 2 (24-hour spaces)	Version 3 (10-minute and 24-hour spaces)
Day 1	12 minutes of chemistry 10-minute 'space' 12 minutes of chemistry repeated 10-minute 'space' 12 minutes of chemistry repeated	12 minutes of chemistry 12 minutes of physics 12 minutes of biology 20 minutes of 'space' at end	12 minutes of chemistry 10 minutes of 'space' 12 minutes of physics 10 minutes of 'space' 12 minutes 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

By the end of this developmental pilot, HTSA had developed SMART Spaces into a programme that could feasibly be delivered in English schools. Both teachers and pupils gave substantial positive feedback about the programme. Most teachers did not feel they needed much further support to deliver the intervention, which suggests the training was successful. Teachers generally reported that they delivered SMART Spaces lessons as prescribed and did not make major alterations. The programme is ready for a larger trial; it is clearly defined and could be delivered to the large number of schools required in an efficacy trial. The evaluator estimated that the programme would cost schools £10 per pupil in the first year of delivery, a very low cost. Schools will also need to arrange one full day of cover, and this might result in further costs.

How was the pilot conducted?

This evaluation was designed to provide feedback to the Hallam team throughout their development of the programme, and, at the end of the project, provide the EEF with a judgement of the programme's evidence of promise, feasibility, and readiness for trial. It was conducted over one and a half years and divided into three phases:

- In the first phase, CESI worked with HTSA to carry out a literature review, which informed the design of a logic model for the programme and selection of three versions of the programme for later testing (see Table 1).
- The second phase used qualitative methods and worked with four schools to examine the feasibility of the three programme variations. The programme was adapted in response to feedback from this feasibility stage.
- The final phase aimed to provide some preliminary evidence about the effectiveness of the three different versions. The three programme variants were compared against a control group which received the PowerPoint slides, but no spacing protocol, and a control group which received neither slides nor spacing protocol.

Table 2: Summary of pilot findings

Question	Finding	Comment
Is there evidence to support the theory of change?	Yes	The spaced learning principle is supported by evidence from both the cognitive science and neuroscience literature. The version which combined ten-minute and 24-hour spaces appeared to be the most promising. However, a larger trial is required before drawing any firm conclusions about the effectiveness of SMART Spaces.
Was the approach feasible?	Yes	The programme was delivered successfully and was acceptable to both teachers and pupils.
Is the approach ready to be evaluated in a trial?	Yes	SMART Spaces is a well-defined and scalable programme that is ready for an efficacy randomised controlled trial (RCT).

Introduction

Intervention

This research report details the design, feasibility, and optimisation of a GCSE science attainment intervention called SMART Spaces*,* constituting an early phase innovation or 'proof of concept' project. The main driver for this research was the mutual interest of the Wellcome Trust and the Education Endowment Foundation in the translation of neuroscientific evidence into teaching practice. The research in this report was informed by neuroscientific evidence and cognitive experimental psychology evidence, as well as educational practice literature in these two research areas.

The appointed evaluation team, the Centre for Evidence and Social Innovation (CESI), and project team, Hallam Teaching School Alliance (HTSA), agreed at an early stage that a research and practice partnership would be the best approach to adopt in order to design, assess feasibility, and optimise the emerging programme. An iterative feedback strategy was then employed by the partners throughout, involving a 'develop, test, feedback, consult, amend, and re-test' loop. The main output of this working partnership, and the research conducted in this study, is the latest version of SMART Spaces, which is described in the TIDieR checklist below in Table 3 (a checklist detailing the training and materials relating to the programme, see Hoffman *et al*., 2014).

Table 3: TIDieR checklist for SMART Spaces programme

Item No.	Item
Brief Name	
1	SMART Spaces: Spaced-learning for Memory And Retention Teaching (GCSE Science Revision Version).
Why	
2	Educational programme for GCSE students primarily used for exam revision with an aim to improve science attainment for Year 9 and 10 pupils.
What	
3	Materials: SMART Materials, PowerPoint slides, SMART Spaces manual, and SMART Spaces activity pack.
4	Procedures: SMART Training—teachers are trained on delivery of SMART Spaces in a one-day training session. SMART Spaces implementation: biology, chemistry and physics topics were taught in the three short 12-minute lessons with 10-minute spaces between each topic; this process was repeated over three consecutive days (thus providing additional 24-hour spaces between content repetitions).
Who provided	
5	SMART Training was provided by trainers experienced in the delivery of SMART Spaces. Future teachers who deliver the programme will be GCSE science teachers who have SMART Training. The same teacher should provide the whole session of SMART Spaces on the three consecutive days.
How	
6	Whole-class programme that is conducted during three normal science lessons.
Where	
7	SMART Training conducted in out-of-school session, and SMART Spaces conducted in standard GCSE classroom.
When and how much	
8	The programme covers GCSE science curriculum content in a high intensity way. The SMART Spaces slides are set out in three 12-minute chunks of GCSE chemistry, physics and biology (approximately one third of each course) content to be taught in one-hour lessons, repeated on three consecutive days.
Tailoring	

9	The programme logic model was designed using neuroscientific evidence, cognitive psychology evidence, and educational practice literature in both areas. Feasibility was piloted in four schools and optimisation was achieved by trialling three different variants against two controls in 12 schools (detailed in this report).
Modifications	
10	The optimisation 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 ten-minute and 24-hour spaces in the delivery of SMART Spaces content. 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 ten-minute and 24-hour spaces.
How well	
11	Planned: effective implementation required training teachers in all twelve schools before they took part in the optimisation trial and all delivered their version of the programme. This training was planned to consist of modelling, practice, and feedback on programme 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 variant there was no apparent difference in implementation. The content of the SMART Spaces programme was found to have a significant benefit over a no spaces/materials control in the optimisation 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: ten-minute and 24-hour intervals, SMART Training, and SMART Resources.

The TIDieR checklist details the two main elements to the SMART Spaces programme—SMART Materials and SMART Training. The SMART Materials comprise a manual and an activities pack. The manual is a comprehensive guide to the SMART Spaces programme and is intended to help teachers deliver the programme with fidelity (that is, in a manner consistent with the original design) in any classroom. At the time of writing, the SMART Spaces manual is under production with a view to facilitating the scaling-up of the intervention in a larger trial. It essentially covers the following elements: background evidence relating to the programme's development (a summary of this report), the programme logic model, the slides for teachers to use during the sessions (chemistry, physics and biology GCSE content—see example in Appendix 4), and a step-by-step guide on how to deliver the programme.

The SMART Training consists of a one-day training course with an experienced trainer in the delivery of the programme (usually a GCSE science teacher). SMART Training is a prerequisite for all delivery teachers. It includes the presentation of some of the supporting evidence from neuroscience and cognitive psychology, but the major component is modelling how the programme is delivered, as well as practice and feedback for the teachers on their delivery of SMART Spaces. Specifically, the training schedule is:

- the scientific background to SMART Spaces—the how and the why of why it works (20 mins);
- how the sessions are managed, including managing the activities in the 'spaces' (15 mins);
- a look at the lesson resources provided (15 mins);
- an experience of how a SMART Spaces session actually runs (20 mins); and
- the opportunity to have a go at delivering a session to the other delegates with constructive feedback (20 mins per teacher).

Finally, the SMART Activities is a pack of materials that are used in the ten-minute ‘spaces’ (including juggling balls), and includes a description of how to conduct the activities in various classroom settings.

Background evidence

The relationship between learning and spaces in time

Since Ebbinghaus (1885) first discovered that he could learn material from fewer repetitions when the repetitions were spaced apart in time rather than grouped together, the relationship between time and learning has fascinated everyone from school teachers to neurobiologists. ‘Spaced learning’ (also often referred to as ‘distributed learning’) is a learning strategy in which two or more study periods are separated in time by an inter-study interval (ISI). The ISI may be as brief as ten minutes, 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 describes an approach where all studying or stimulus presentation occurs in one constant block, without any gaps in time to separate periods of learning. The ‘spacing effect’ refers to the benefit of *spaced learning* on memory and retention of information over *massed learning*.

Evidence (reviewed below) suggests the same material can be learned more effectively using a spaced learning strategy rather than a massed learning strategy and, assuming that the overall quantity of studying is equal in time between the two approaches, this is called a spacing effect. This should be distinguished from the concept of a ‘lag effect’ which describes the benefit of one length of ISI over another for learning, when two or more spaced strategies are compared. For example, if the same material was learned more effectively with a 60-minute ISI between study periods than with a ten-minute ISI, this would be referred to as a lag effect. The comparison of different ISIs has been used to attempt to determine the optimal spacing strategy for different retention intervals (RI). The retention interval is the time between the final study period in a spaced learning procedure—or the single massed study period in a massed learning procedure—and the testing period.

Another related concept to spaced learning is interleaving. In a situation where material must be learnt about different topics, interleaving these areas of learning, rather than blocking them together, may lead to better retention. For example, interleaving may follow a pattern of ‘ABCABCABC’ rather than an ‘AAABBBCCC’ blocked pattern of delivery. Taylor and Rohrer (2010) used a combined approach of spacing and interleaving for the practice of four types of mathematics problem-solving with children. They found that, when spacing was controlled between blocked and interleaved practice, interleaved practice led to higher test scores after the practice, despite the fact that performance on the practice day was impaired. It may be the case that interleaving leads to a more difficult practice experience in the short term, but improves retention in the mid to long term.

What has past research from cognitive psychology and neuroscience told us about spaced learning?

Cognitive psychology and spaced learning

Numerous reviews of cognitive experiments on spacing effects have revealed the robustness of spacing effects for simple memory tasks and motor skills. These reviews also opened up the possibility that there is an optimal spacing interval which depends on the retention interval and the complexity of the task. It should be noted that, unless otherwise specified, these cognitive psychology reviews typically focus on adults. Lee and Genovese (1988) conducted a literature review of abstracts on distributed learning of motor skills published since 1968 in *Psychological Abstracts*, and prior to 1968 in *Perceptual and Motor Skills* (1949 to 1968). They reviewed 47 articles and found that spaced learning improved both acquisition and retention of motor skills (mean ES = 0.91). Moss (1996) reviewed 120 articles on the spaced learning effect, comparing various types of learning material (verbal information, intellectual skill, and motor learning). Longer spacing intervals improved the learning of verbal information and motor skills in over 80% of the studies reviewed. However, for more complex intellectual tasks (such as mathematical computation), only one third of the studies showed a spacing effect.

Donovan and Radosevich (1999) reviewed 63 articles, including both experimental and quasi-experimental studies (mean weighted ES = 0.46) and found that the largest spacing effects were found in studies of learning low-complexity tasks (such as typing). They also found that, for verbal tasks, extending the space improved recall up to a point, and then began to negatively affect recall if the space was extended too much. Janiszewski *et al.* (2003) reviewed 97 articles on spacing effects and space lengths for various types of task. The authors included studies on the basis that they reported a spacing effect, and included enough statistical information to calculate an effect size. They found that numerous factors resulted in larger spacing effect sizes including: longer spaces (mean ES = 0.57), complexity of stimuli (mean ES = 0.42), less meaningful stimuli (mean ES = 0.51), and intentional (rather than incidental) learning (mean ES = 0.35). Cepeda *et al*. (2006) reviewed 317 spacing effect experiments in 184 articles, including children, adults and older adults. These were all verbal memory learning tasks, followed by a recall test. The authors included both an analysis of spaced versus massed learning, and a comparison of different lengths of spacing intervals. All but 12 of the 217 studies showed a benefit of spaced learning over massed study, but it is important to note these studies reported only improvements in accuracy (due to insufficient effect size data). Cepeda *et al.* also found that increasing the spacing interval increased recall, but too long a spacing interval for a given retention period reduced recall (in agreement with Donovan and Radosevich, 1999). Cepeda *et al*. determined that the spacing interval during studying should increase as retention interval increases to optimise recall. When the retention interval was less than one minute, optimal space was also less than one minute. At least a one-month space is necessary for optimal recall after six months. This was further explored by Cepeda *et al.* (2009), who asked 215 undergraduates in the U.S.A. to learn Swahili-English pairs in one experiment, and pairs of facts and objects in another. Participants were randomly assigned to one of six groups, each group being assigned a different ISI, varying from five minutes to 14 days for the first experiment, and 20 minutes to six months for the second. They also examined retention intervals of ten days and six months. It was found that increases in the spacing interval led to a sharp improvement in retention, but if the spacing interval became too long, a slight decrease in retention was observed. For a six-month retention interval, it was found that a 28-day space was optimal.

The body of cognitive psychology literature provides strong evidence for the benefits of spaced learning and for the model of an optimal ISI for a given retention interval in which the ISI is appropriately long in relation to the length of the RI. At this point we should note that what happens *during* the space may also influence the success of the learning. Perhaps most pertinent is the role of sleep. Bell *et al.* (2014) asked participants to learn Swahili-English word pairs and compared massed learning versus three types of spaced learning (12-hour space without sleep versus 12-hour space with sleep versus 24-hour space). Participants (n = 141) were randomly assigned to one of the four conditions. A spacing effect was found for all the spacing conditions, but interestingly, the two conditions with sleep (12 hours with sleep and 24 hours) showed a stronger advantage over massed learning than the 12 hours without sleep condition. It is possible that longer ISIs themselves are not solely responsible for their benefits over shorter ISIs, and that the sleep associated with longer intervals plays a role.

Overall, it is apparent from the reviews of cognitive psychology experiments that spacing effects can be consistently found, that long spaces of a day or more are increasingly beneficial for learning, and that sleep during the space may be important for the success of memory formation. These three facets of the evidence provide a solid base for the design of a 24-hour space approach to spaced learning.

Neuroscience and spaced learning

Studies within the field of neuroscience have explored the neurobiological basis of the spacing effect. Some of these have involved removing cells from animal brains (post-mortem) and stimulating them with either neurotransmitter chemicals or electricity. This is to mimic the input of stimulation from learning material to a live brain. Then the response of the brain cells is recorded, as a proxy of the sort of response that might happen during actual memory formation.

These studies have focused on recording neurobiological indicators of 'long term potentiation' (LTP)—a process through which synapses may become stronger after being stimulated, and thus transmit a long lasting signal between neurons. Hebbian theory states that cells which are stimulated together form a long lasting connection and this is thought to be one mechanism for how memories are formed in the brain. This theory is often abridged to 'cells which fire together, wire together'. LTP is often investigated as a potential biological basis for the formation of long-term memories. Electrical activity of synapses, proteins, and gene presence can be recorded in a cell as an indicator of LTP, and therefore as a potential precursor to the formation of long-term memory.

Mauelshagen, Sherff and Carew (1998) exposed synapses removed from Aplysia (marine molluscs) to serotonin in either five bursts of five minutes with 15-minute intervals, or one long massed exposure of 25 minutes. They found that electrical responses to stimulation 24 hours 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 studies where synapses were taken from sections of rat hippocampi. The hippocampus is where consolidation from short-term to long-term memory is thought to take place. Fields and his colleagues studied electrical activity and the presence of genes in these rat brain cells after electrical stimulation, as a measure of long-term potentiation. Fields found that the rat synapses produced twice the voltage of electrical activity after being stimulated in three bursts with ten-minute spaces than when they were stimulated in a massed pattern. They also found the presence of the protein CREB and a gene called zif268 after spaced stimulation, both of which we know are associated with the conversion of short- to long-term memory. These studies demonstrate the neurobiological processes involved in the cognitive advantages of spaced learning in terms of electrical activity, proteins, and genetic indicators of memory formation.

As seen in the review of the cognitive psychology literature, the neuroscience literature also points to the benefit of increasing the space length for better retention. Kramar *et al*. (2012) used a similar experimental design to Fields (2009), stimulating rat brain cells with electricity but with the addition of exploring the effects of different spacing intervals—10, 30 and 60 minutes between periods of stimulation. Twice as much long-term potentiation was produced when a 60-minute interval was used rather than 10 or 30 minutes. Zhang *et al*. (2011) found that a spacing protocol including a 30-minute space in the stimulation of mollusc brain cells led to greater activity than the shorter spaces used in the Mauelshagen *et al.* (1998) study. Overall, the animal brain studies reveal that there is physical evidence of long-term memory formation being enhanced by spaced stimulation, and more recently that longer intervals than originally suggested by Fields (2009) may be optimal.

Much less information is available about human brain activity during spaced learning, although some new developments are beginning to emerge. Earlier neuroscience research with humans did not actually involve explicit spaced learning tasks. Van Strien *et al*. (2007) had 22 adults judge if they had heard a word before while being presented with a list of repeated words. Recall was better for words which had been presented with spaced repetitions than after the massed repetitions. During the task, electroencephalography (EEG) recordings were made. EEG involves the recording of electrical activity at the scalp, which allows the size of event-related potentials (ERP, measured in voltage) to be used as an index of the brain's response. Van Strien *et al.* (2007) found a larger change in event-related potentials associated with memory search and template matching (N400 and LPC) in response to massed rather than spaced presentations. This means that a pattern in electrical activity which may be indicative of memory activity showed evidence of increased difficulty performing a memory task when massed presentation was used rather than spaced presentation. This suggests that massed presentations 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 (2011) compared the paired learning of nouns and pictures from two repetitions presented in either a massed or spaced format (12-second ISIs). They found that massed learning resulted in weaker brain responses when the items were presented the second time in comparison with the first time. Spaced learning did not show this suppression of brain responses. 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.* (2012) asked adults to memorise 120 novel faces, some of which were repeated in a massed manner, and some in a spaced manner. Spaced learning led to more activity in a brain area associated with face recognition (bilateral fusiform gyrus) than massed learning. Spaced learning also led to better learning of items than massed learning. Again, this indicates that spaced learning leads to less suppression of learning responses for repetitions than massed learning.

The human-based neuroscience studies help explain the benefits of spaced learning in terms of increased attention to repeated items. Due to the highly technical and time-consuming nature of human neuroscience studies, we have not yet seen the recording of human brain responses to long spacing intervals. The early neuroscience work with animals showed spacing effects with ten-minute spaces originally, but there is emerging evidence of the benefits of longer spaces over short.

Spaced learning in educational settings

Although the reviews summarised above show that the spacing effect is highly robust, some of the results suggest that for more complex tasks, spacing may provide less of a benefit than it does for simple tasks. However, more recently, progress has been made in determining that spacing effects can be observed even for the complex learning in normal educational practice. For example, Miles (2014) applied the evidence of the spacing effect from cognitive psychology in the classroom and examined the effect of spacing on the teaching of English grammar to South Korean undergraduate TESOL students, randomly assigning students to massed or spaced conditions in an experimental study. He compared massed against spaced learning (ISIs of one week, then four weeks, with a retention interval of five weeks). This study found a strong spacing effect, despite the complex nature of the learning task. Again, we see the benefits of longer intervals between study periods for a long retention interval.

This benefit of long inter-study intervals for long retention intervals is apparent in numerous quasi-experimental studies of spaced practice in education. Carpenter *et al.* (2009) conducted learning sessions during primary school class time. Eighth-grade children in the U.S.A. were given a history test, then a review of answers and facts. This review was then repeated either one week or 16 weeks later, followed by a retention test nine months after the review. Children with the ISI of 16 weeks recalled more facts after nine months than the one-week ISI group. Bird (2011) compared ISIs of three and 14 days for spaced learning lessons of grammar for English as a second language to university students. After a retention interval of seven days, no difference between the two ISIs was found, but for a 60-day retention interval, only the 14-day ISI group showed sustained improvement in comparison with a pre-test. The three-day ISI group actually significantly decreased in performance between the seven-day test and the 60-day test, whereas the 14-day ISI group showed a score consistent with their seven-day score. Psychology undergraduates in Canada taught using a spaced protocol with eight-day ISIs showed better retention after five weeks than when one-day ISIs were used (Kapler, Weston and Wiseheart, 2015). The eight-day ISI group did not differ from the one-day ISI group on an initial test that occurred before the learning period, yet the eight-day ISI group scored significantly higher than the one-day ISI group on the final retention test, with the eight-day ISI group scoring 15% and 20% more highly than the one-day group on both simple and complex question types respectively.

There has also been an attempt to apply the evidence of a spacing effect on short ISIs from the neuroscience literature. In 2013, Kelley and Whatson successfully employed spaced learning strategies in the classroom with children aged 14 to 15 years in England, in a quasi-experimental study. Teachers were trained to use spaced learning techniques, and assessment was a high stakes test on a national curriculum biology course. They claim that 90 minutes of spaced learning (three periods of 20 minutes of teaching, interspersed with ten-minute intervals of distractor activities) produced retention of the information that was not significantly different to four months of typical teaching (massed learning), despite significantly less teaching time. This clearly demonstrates how a spaced approach can be considerably more efficient than a massed approach. Kelley and Whatson (2013) designed their spacing protocol to mirror the neuroscience work of Fields (2009) who used ten-minute ISIs to successfully produce increased long-term potentiation in rat brains.

Although this is a good example of neuroscientific evidence-based design for educational practice, it should be noted that the scientific evidence, from both neuroscience and cognitive psychology, is now pointing towards intervals of longer than ten minutes being more beneficial (Kramar *et al*., 2012), and especially so when the retention interval is longer than a few hours (Cepeda *et al*., 2009; Miles, 2014).

Interpretation of the literature

Research in an educational setting (Kelley and Whatson, 2013) suggested that ten-minute ISIs were of benefit, as did the rat hippocampal stimulation work by Fields (2009). Systematic reviews (detailed in Appendix 1) suggest that longer ISIs are needed for more practical education benefits. Cepeda *et al.* (2006) found that for retention intervals of a month or longer, an ISI of at least one day is needed. Later rat hippocampal stimulation research—similar to the work of Fields (Kramar *et al.* 2012)—suggests that 60 minutes is optimal for long-term potentiation, rather than shorter intervals of ten minutes. Longer ISIs have been found to be of benefit when long retention intervals with complex material are used (Miles, 2014), even in a classroom setting (Kapler *et al.,* 2015).

Significant benefits for spacing of synaptic stimulation are similar to behavioural learning (Fields, 2009; Kramar *et al*., 2012). However, Kornmeier and Sosic-Vasic (2012) comment that we do not know if this long-term potentiation facilitation seen in the lab-based animal studies actually contributes to the complex pattern of behavioural benefits (for example, non-monotonic lag effects described by Cepeda *et al.*, 2006). This means that we do not know the extent to which these changes in brain cells seen in the lab experiments are actually contributing to the real-world changes in memory we see in more practical spaced learning studies. For example, these changes in long-term potentiation seen in the lab do not offer much explanation of the mechanisms at play when we see that memory retention increases up to a certain point when spacing time is increased, but then begins to tail off.

The vast majority of neuroscience work on spacing effects is very far removed from the classroom. EEG research could bridge the gap as it involves the recording of brain activity both during the presentation of stimuli and during recall, but there is a need for more focus on the efficacy of longer inter-study intervals to determine optimal conditions for meaningful retention periods. Non-uniform (Zhang *et al.,* 2011) or adaptive protocols (Kerfoot, 2010) for spacing may be optimal on the individual level, but less applicable in classrooms. Non-uniform spacing protocols involve a range of different duration spaces within a participant’s learning experience, and adaptive protocols tailor the number of repetitions to the individual, dependent on performance. These types of spacing may offer performance or time benefits over regular spaces and predetermined repetition numbers, however they would likely not be feasible for classroom delivery due to the high demand of assessing each participant’s learning throughout the process.

Many hypotheses have been provided for the causal driver of the spacing effect—from protein synthesis, electrical activity, through to sleep (a more detailed summary of the research literature is available in Appendix 1 and from Gerbier and Toppino, 2015). The cognitive psychology and educational practice literature tells us about the mechanisms and benefits of longer spaces. The neuroscience literature is focused on shorter spaces. Yet it is important to consider that cognitive changes are still at play with shorter spaces, and neurological changes are still at play with longer spaces—these pairings have just not been the focus of the respective bodies of literature. However, the focus of the current report is to explore the more practical questions around optimum spacing length and feasibility in the classroom.

Overall, we can see that there is evidence in both the neuroscientific and cognitive psychology areas of research for a robust spacing effect. However, due to recent progress in both the experimental research of cognitive psychology and research exploring spaced learning in an educational context, we can see that there are two distinct sources of evidence for short ISIs (from neuroscientific literature) and for longer length retention intervals (from cognitive experimental literature). In addition, there is limited evidence about the feasibility of using spaced learning in the classroom in both literatures and none comparing the two lengths of spaces. In fact, Dempster (1988) has called for more applied research on the duration and type of spaces which may be beneficial to learning.

With this evidence from the research literature in mind, the research and project team co-designed a logic model to compare different ISI periods (ten minutes versus 24 hours). Furthermore, we set up a further two phases to this investigation. The first phase aimed to test the feasibility of several variants of spaced learning with different ISIs in the classroom. The second phase was a small pilot controlled trial with the purpose of identifying the optimum variant of spaced learning for improving attainment, which would ultimately characterise the SMART Spaces programme.

Logic model and research rationale

This section outlines the logic model underpinning SMART Spaces and the rationale for the research design utilised in this study (see Figure 1). As an innovation project, this study had three stages exploring all the elements of a standard programme design logic model—(1) theoretical development (exploring programme inputs), (2) feasibility (investigating programme outputs), and (3) optimisation (testing programme outcomes). Theoretical development was conducted using the research literature and consultation between the evaluation and project teams. The feasibility study (FS) examined the implementation of spaced learning in four schools and trialled a potential science outcome measure (GCSE attainment). The optimisation study (OS) was an experiment with a pre-test and post-test of GCSE science questions (using the bespoke measure selected from GCSE questions that was piloted in the FS) in 12 schools. Three variants of the programme were tested in the OS, with the addition of two control groups in order to gain some early insight into the relative efficacy of the programme variants. In addition to the outcome assessment, the OS also included a reliable measure of pupil engagement (developed by the evaluation team) to investigate the influence of implementation on outcomes.

The study was designed in this way to provide iterative feedback to improve the innovation and design of the programme as well as to inform the research design at each stage and produce some indication of the promise of the programme improving science attainment. There is one important caveat to be remembered about this investigation: this study was NOT set up to be a full efficacy trial of SMART Spaces (that is, having an appropriately powered sample size and requisite number of schools); rather the purpose was to develop a spaced learning programme, to provide evidence for the feasibility of the programme and to optimise it for a future study of its efficacy.

The SMART Spaces logic model was generated from the evidence provided by the different bodies of literature on spaced learning, namely neuroscience and cognitive psychology literature. This resulted in the creation of several variants of spaced learning instruction to be tested, each variant mapping onto a theoretical framework provided by the literature. These variants were all designed to be taught within a one-hour lesson per day over three consecutive days, with the same amount of time spent on teaching the subject material. The lessons were designed to be an hour long so that they would fit the majority of school timetables. The only difference between the variants was the length and distribution of spaces. These variants were tested in the context of revision lessons for GCSE science, not for the initial learning of the science content. The following paragraphs detail the rationale behind each of the variants and controls that appear in the logic model.

Variant 1 (ten-minute variant)

The neuroscientific literature provides evidence of increased memory markers when ten-minute or similarly short spaces are used between bursts of stimulation, in comparison with massed stimulation (for example, Fields *et al*., 2009; Menzel *et al*., 2001). This is explored through Variant 1 (called the 'ten-minute variant'). In practice, the pupils receive 12 minutes of 'Subject A' (for example, chemistry), followed by ten minutes of space activity (typically juggling), a further 12 minutes repeating Subject A material (the same slides as before), another ten-minute space activity, then a final 12 minutes of Subject A. This exact pattern is repeated on day two but with a change to the material provided (Subject B, such as physics) and repeated again on day three with material changed once more (Subject C, such as biology).

Variant 2 (24-hour variant)

Experimental cognitive psychology literature provides evidence that longer spaces (24 hours or more) may be optimal when long retention periods are required, such as weeks or months (for example, Bell *et al*., 2014; Carpenter *et al.,* 2009; Cepeda *et al.,* 2009). This is partially explored, with medium-term spaces, through Variant 2 (called the '24 hour variant'). In practice, the pupils receive 12 minutes of Subject A, followed by 12 minutes of Subject B, then 12 minutes of Subject C followed by 20 minutes of space activity (to ensure a consistent pattern of 36 minutes content is presented with all variants). The same pattern is repeated on day two and day three. The interval may not be exactly 24 hours—simply the next day.

Variant 3 (ten-minute and 24-hour variant)

Both of these broad approaches ('ten-minute spaces' and '24-hour spaces') may contribute to the short-term outcome of memory formation and subsequently to the later outcome of educational attainment through both similar and different mechanisms of change. To explore this, the research was designed to test a third variant, Variant 3, which combined both 'ten-minute' and '24-hour' spaces in the delivery of materials. In practice, the pupils receive 12 minutes of Subject A, followed by 10 minutes of space activity, then 12 minutes of Subject B, another 10 minutes of space activity, and finally 12 minutes of Subject C. The exact same pattern is repeated on days two and three.

Control 1 (slides only control)

There are also two controls in this study. Control 1 is a 'slides only' control, which provides the PowerPoint slides of teaching content in the normal time, with the inter-study task (for example, juggling) performed after all content is provided, and thus providing no spaces in the learning. This is to assess for any effect that the condensed study slides may have on the learners' attainment outcomes. In practice, the pupils receive 12 minutes of Subject A, followed by 12 minutes of Subject A, then 12 minutes of Subject A, followed by 20 minutes of space activity (that is, both ten-minute spaces combined at the end). The same pattern is repeated on day two changing the content to Subject B, and day three changing the content to Subject C.

Control 2 (no spaces/materials control)

The second control is a pure control, which provided no slides or spaced learning to the pupils and they proceed with their normal science lessons. However, they were pre- and post-tested like all the other pupils engaged in the variants and control conditions described above.

In summary, it is important to consider that there is evidence in the literature for *both* ten-minute and 24-hour spaces being valuable for memory formation. This study will therefore examine the contribution of each of these ISIs, as well as a combination of the two, on science attainment.

As has been discussed, there are a number of theories of change behind the spacing effect, many of which have been explored in the literature (Gerbier and Toppino, 2015). However, it is suggested that any observed cognitive learning and neurological changes are not specific to long and short spaces respectively. Neurological changes might still occur as a result of longer spacing, they just have not been explored in the neuroscience literature. Similarly, cognitive changes might still occur as a result of shorter spacing, they are just not the focus in the cognitive literature. The logic model (Figure 1) does highlight these links (dashed line) between cognitive and neuroscientific changes as a result of spacing effects. Regarding the theory behind the SMART Spaces programme, this study does not explicate a theory of change, rather it explores the more practical theory of intervention behind spaced learning, that is, on what length of spaces are optimal for producing positive effects on memory and learning, and whether it can be delivered in a feasible way in a classroom.

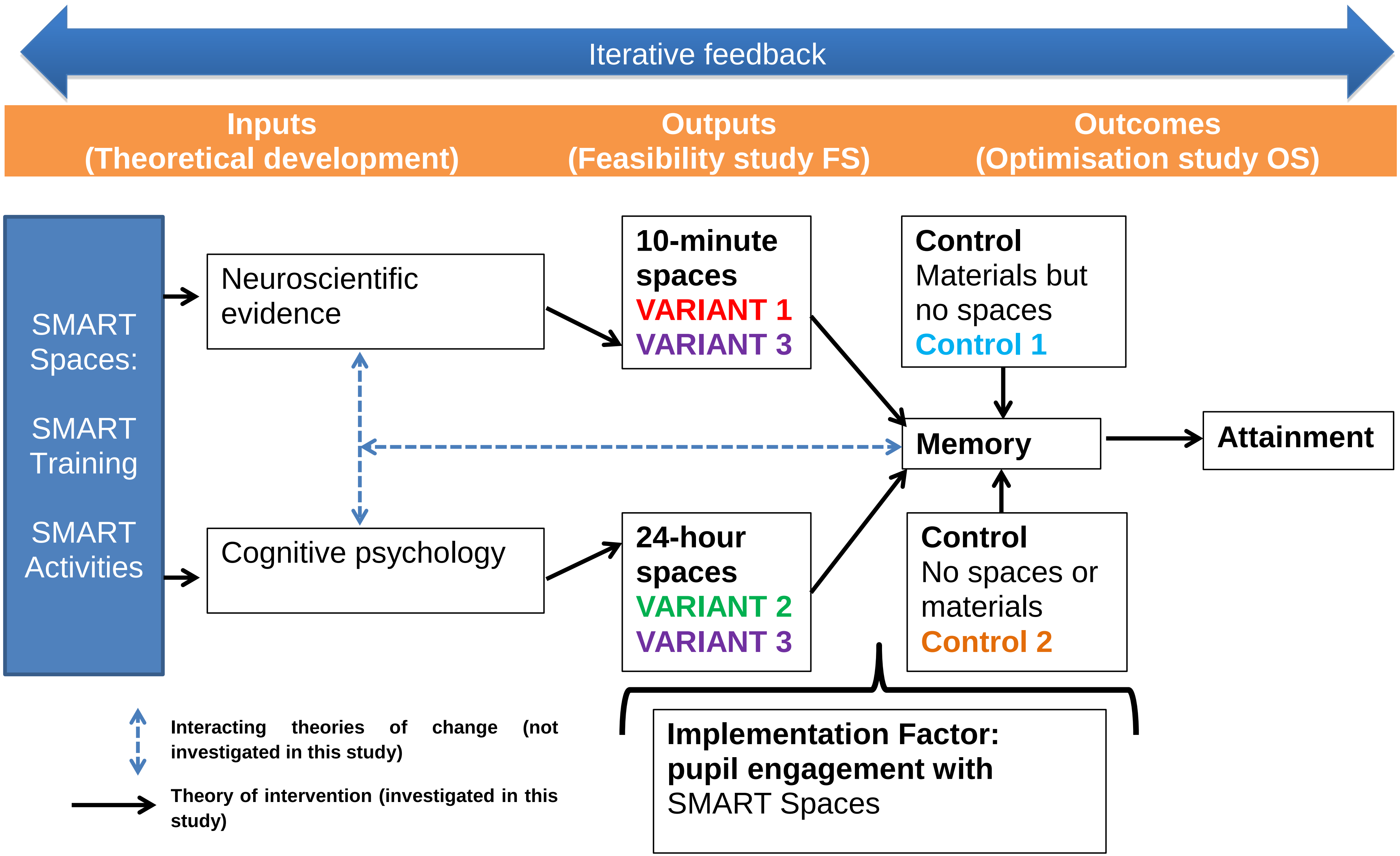
Lastly, it is well understood that educational programmes do not operate in a fully controlled environment. Therefore, it is important to understand contextual factors and their influence on programme effectiveness (Bonell, *et al.,* 2012; Craig, *et al,* 2008; Jamal, *et al*., 2015). To do this we explored the influence of implementation factors on programme effects. One major issue explored was pupils' engagement with the spaced learning—using a standardised post-test measure—as well as the type of spaced activity (was it juggling or some other activity?) so we could relate these factors to any observed outcome change.

Research questions

With the logic model and rationale in mind, we considered the following research questions:

1. What can we learn about the feasibility of implementing a 'spaced learning' programme in a practical classroom situation?
2. What length of spaces showed the best evidence of promise on attainment outcomes? Was it ten-minute spaces, 24-hour spaces, or a combination of the two?
3. What implementation factors have an influence on programme effects?
4. Is SMART Spaces feasible enough, and with enough evidence of promise, to warrant a future efficacy trial?

Figure 1: SMART Spaces Logic Model and Research Rationale



Ethical review

All research was conducted according to QUB School of Education ethical guidelines. Ethical consent was obtained from the Ethics Committee before data collection was conducted. Informed consent was obtained from participant schools and all pupils had a right not to complete or withdraw their data. The data collected was coded and entered onto a database, made anonymous, and held securely on a password-protected computer. No NPD data was required for this project.

Project team

Centre for Evidence and Social Innovation—Evaluation Team

- Dr Liam O'Hare (Principal Investigator)
- Dr Andy Biggart
- Professor Carol McGuinness
- Dr Patrick Stark
- Professor Allen Thurston

Hallam Teaching School Alliance—Project Team

- Alastair Gittner (Project Team Leader)
- Farhana Zaman (Project Manager and Administrator)
- Steve Davis (Head of Hallam TSA)

Methods

Recruitment

The schools were recruited in both study phases (feasibility study and optimisation study) by the project team.[1] Four schools were recruited for the FS, and 12 for the OS. The four schools in the FS had already been using a version of spaced learning (predominately a ten-minute variant) for a few years and had a range of familiarity with the concept. The 12 schools in the OS had not previously implemented spaced learning. For these schools, recruitment advertisements were shared across England by the implementation team who then responded to interested schools with further information. Only 12 schools agreed to take part. There were no selection criteria for the 12 schools in this part of the study beyond an expression of interest, and none were excluded from the study. However, 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 for disadvantage).

Consent was sought at the school level through a memorandum of understanding (see Appendix 3). The 12 schools in the OS were randomised by the evaluation team into the five conditions:

- Variant 1 (ten-minute version): three schools;
- Variant 2 (24-hour version): two schools;
- Variant 3 (combined ten-minute and 24-hour version): three schools;
- Control 1 ('slides-only' control): two schools; and
- Control 2 ('business as usual' control): two schools.

All GCSE science pupils in the school were eligible for the programme on the condition that their class teachers had received the SMART Spaces training. Classes of pupils were chosen within each participating school by the Spaced Learning contact for each school. The 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. Consent was sought at the pupil level through opt-out consent forms (sent home to parents and verbally explained to pupils at testing) for participation in the pre-tests and post-tests.

Data collection

1. Theoretical development (months 1–6)

In order to gain insight into potential theories to underpin the SMART Spaces programme, a structured literature review of 'spaced learning' interventions was conducted covering the elements of a standard logic model—inputs (resources), outputs (activities) and outcomes (changes), assumptions (relating to training, quality of delivery, engagement, and so on), and external factors (current 'spaced learning' practice and proliferation as well as the educational context). The literature review also examined the underpinning theory around spaced learning in terms of a theory of change and a theory of intervention. This literature review covered multiple perspectives on spaced learning research and included neuroscientific, cognitive psychological, and educational practice literature.

The literature review fed into a service design and logic modelling process. This process was a collaboration between a range of stakeholders: the evaluation team (QUB), project team (Hallam TSA), school principals, and teachers. The process began with a programme development retreat that

[1] Recruitment for the OS was the responsibility of the project team, with support and assistance from the evaluators. All schools recruited were English State schools. Schools had varying degrees of familiarity with the concept of spaced learning but none was implementing it.

included capacity-building in logic modelling, presentations on spaced learning literature, and current spaced learning practice.

The logic model produced during the retreat informed the design of the spaced learning programme (see Figure 1). Three variants of spaced learning and two control conditions were designed as a result of this theoretical development process. The evidence from the early neuroscience literature informed the development of the ten-minute space condition. Evidence from cognitive psychology, more recent neuroscience literature, and educational practice was the basis of the 24-hour space condition. A combined approach of using both ten-minute and 24-hour spaces was also designed (with the assumption it could bring together the impact of both short-term and medium-term mechanisms of change). A control condition, using the spaced learning teaching materials but without the spaces between learning periods, was designed to test for the effect of the materials alone. This design was explored with a group of teachers and the project team to develop activities, PowerPoint resources, training, and lesson plans for the FS pilot study.

2. Feasibility study (months 7–12)

Having produced a draft manual in the first phase, the feasibility study explored the implementation of the programme (with its three variants and one control) in practice. The project team worked closely with a small number of schools in this study (n = 4).[2] Classes in each participating school were assigned to the three different variants of spaced learning or to the control condition. Eight classes participated in this study: four classes were assigned to the ten-minute variant, one to the 24-hour variant, two to the combined ten-minute and 24-hour variant, and one to the 'no spaces' control condition. All schools had one class implementing the ten-minute variant—as they were familiar with its use—but also agreed to try one of the other variants as well (see Table 4).

Table 4: Four feasibility study schools and their variants piloted

School	Class 1 Variant	Class 2 Variant
1	ten-minute	ten-minute/24-hour
2	ten-minute	24-hour
3	ten-minute	no spaces control*
4	ten-minute	ten-minute/24-hour

*i.e. using the spaced learning teaching materials but without the spaces between learning periods.

To initiate the FS, the teachers met with the evaluation and project teams. The group of teachers implementing the programme in the FS had familiarity with the ten-minute programme and agreed to deliver it or train their colleagues when required to provide the programme. No teachers were trained in the other programme variants before the FS was conducted. However, it was explained to them how to change their processes to deliver the other variants during the meeting.

Five pupils in each school took part in focus groups (20 pupils in total across the four schools) to explore engagement, enjoyment, and classroom practicality of each variant. One teacher or headteacher from each school (n = 4) was interviewed at this stage to investigate engagement with the research project, implementation of the lessons, training, and support procedures.

In addition, after implementing the spaced learning programme, pupils took part in a pilot of the outcome measure that was to be used in the next study to check its usability. This test was developed and administered by the evaluation team. This was to explore the appropriateness of the bespoke science measure, rather than to examine quantitative differences between variants at this early stage.

[2] This was an increase from the three schools stated in the protocol because there were four schools engaged and interested at this stage.

3. Optimisation study (months 13–17)

The next phase was the design of the optimisation study involving 12 schools. Schools were randomised at the school level into one of the three intervention conditions or one of the two control conditions—these were the four conditions studied in the FS plus an additional ‘business as usual’ control (see page 18). Feedback from the FS highlighted the need for teachers who were delivering the programme to be well trained, and thus all teachers received training on spaced learning, including additional practice and feedback on the specific variant that they were being asked to implement, as well as being provided with the appropriate materials. Pupil data was collected quantitatively through the pre-test and post-test science attainment outcome measure piloted in the FS. The test was administered by the evaluation team (the evaluation team were aware of the variant the school was testing, but the project team did not know the content of the test). The pupils also received a post-test implementation questionnaire to investigate implementation, fidelity, and pupil engagement with the programme (see Appendix 2). One teacher from each school who delivered the lessons from each variant (total n = 4) was also interviewed to explore implementation factors such as fidelity, training, and sustainability.

4. Review and final manual production (months 18–21)

The final phase of the project was ongoing with the qualitative and quantitative findings feeding back into the logic modelling process (with the aim of revising it to produce the final programme logic model—see discussion) and the production of this report. Final refinements are being made to the programme design in a series of meetings between the evaluation and project teams. These discussions will produce the final output in the form of a ‘spaced learning’ manual featuring a literature review, logic model, implementation data, a training guide, and programme content. This manual will be available for use in any future effectiveness study of the intervention.

Randomisation

For the OS, randomisation 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 ten-minute and 24-hour variants respectively (to ensure some numbers in these two variants in case of school withdrawal). Four schools were pre-tested after randomisation. Randomisation took place on 23 February 2016; four schools were pre-tested in a window two weeks after that, up to 8 March 2016.

One school (School L, a small independent school) assigned to the 24-hour space variant did not wish to take part in the training and delivery of the programme but agreed to the pre-and post-test (with 11 of the 14 pupils providing complete data). Therefore, it was reassigned to the ‘no spaces/materials’ control group. The study cannot, therefore, claim to be a fully randomised trial. However, as design, feasibility, and optimisation were the foci of the study showing evidence of programme promise rather than actual efficacy of the programme, it was deemed appropriate by the evaluation team to include this school in the analysis.

Measures

Outcome measures were conducted after implementation of the programme in the FS, and at both pre-implementation and post-implementation time points in the OS. The same outcome measure was used on all three occasions—a bespoke science measure, comprising past-paper questions from the

AQA GCSE[3] curriculum. This test comprised two sections: Section A—short answers and multiple choice, and Section B—extended answers. There were 39 marks available for Section A and 18 marks for Section B—a total maximum score of 57. Appendix 4 gives full details of the GCSE topics for each science subject (biology, chemistry, and physics) for each section in the paper, the marks allocated to each topic, and the source GCSE science paper. The short answer and multiple choice section required participants to give answers ranging from one word to two or three lines; the extended answer section required considerably more detail per answer, requiring five to six key points of information. The test had a time limit of 45 minutes in the optimisation study, which was determined by the average time taken to complete the test in the feasibility study. The project team were blind to the content of the outcome measure so as not to influence the content of the training sessions or slides to include additional emphasis on the exam questions used. The reliability of the six scores (Sections A and B for the three subjects) used to calculate the total score on the test had a Cronbach's alpha = 0.88. Pre-tests occurred between 23 February and 13 April 2016; post-tests between 26 April and 20 May 2016. Typically, the period between pre- and post-tests was 62 to 65 days with the exception of two schools where it was 49 and 37 days. The period between pre-test and post-test for each school is shown in Appendix 6. Gain scores (post-test minus pre-test scores) were used in the analysis of the data collected and these are summarised in Appendix 8.

Data from the focus groups provided during the FS was used to design an implementation questionnaire which was administered to pupils post-implementation in the OS (see Appendix 2 for summary statistics of responses). This questionnaire also included a 'pupil engagement in spaced learning' scale which was used in outcome-related analysis (see Table 9). There were 13 items in this engagement scale (see Appendix 2 for more details). Reliability of this measure was very good (Cronbach's alpha = 0.91).

Sample size[4]

As this project was largely a design, feasibility, and optimisation process, sample size was not based on required numbers to identify effectiveness. Rather, sample size was chosen to best suit the feasibility and optimisation testing process. Four schools initially implemented a draft version of the programme and its variants. Twelve schools participated in the OS stage of the research. Table 5 below shows the number of participant pupils in each of the 12 schools, broken down into the variants that they received.

[3] 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.

[4] No power calculation was conducted in this project because it was a design, feasibility, and optimisation study of a programme and not a trial of efficacy. A power calculation would have suggested the recruitment of many more schools than actually participated in this study in order to produce findings with a high degree of power/security in terms of programme efficacy. The findings in this study are very low in efficacy security but very useful in terms of design, feasibility, and optimisation.

Table 5: Variant by school and pupil numbers

Variant	Pupil numbers in schools A-L												Total
	A	B	C	D	E	F	G	H	I	J	K	L	
10-min	0	42	0	0	0	0	0	0	47	21	0	0	110
24-hr	0	0	0	0	0	0	0	44	0	0	31	0	75
10-min and 24-hr	0	0	21	0	43	27	0	0	0	0	0	0	91
no spaces control	39	0	0	0	0	0	40	0	0	0	0	0	79
no spaces or materials control	0	0	0	42	0	0	0	0	0	0	0	11*	53
Total	39	42	21	42	43	27	40	44	47	21	31	11	408

*Reassigned from 24-hour version.

Analysis

Analysis was conducted after each of the two stages, the FS and the OS. After the FS, the pupil focus groups and interviews fed into the logic modelling and research design retreat at the end of the FS. Furthermore, reliability and validity tests of the bespoke measures were conducted to explore whether the measures were valid, acceptable, intelligible, collectable, and capable of showing plausible positive changes.

At the OS phase there was an analysis of the effects of each variant on pre-post outcome measures compared to both control groups. These were calculated using mean gain scores, pre and post SDs, and paired t-tests. Variant scores were compared to control scores on total score short-answer and long-answer questions.
The analysis also explored the relationship between implementation factors (engagement, space activity type, and retention) and outcome change. No clustering was accounted for in these regression analyses as there were five arms to the optimisation experiment and only two or three schools in the five arms. However, multilevel models were conducted and are included in Appendix 7 to provide the reader with this information if required.

Costs

The following cost estimates were developed during the creation of the SMART Spaces manual:

Cost item	Per class (£)
Cost of one class manual	100
Cost of juggling balls for one class	10
Cost of trainers time/number of teachers at training (£450/5)	90
Cost per class in first year	200
Cost per pupil (20 pupils per class) in first year	10

Trainers were paid on a per-hour basis, (around £50 per hour + travel expenses with an understanding that preparation time was needed for each session).

Teachers are required to attend one full day of training, and cover will need to be arranged for this.

Timeline

Table 6: Timeline of project activities

Date	Phase	Activity
2015		
Jan	Set-up	• Ethics application QUB School of Education • Collect information for literature review • Start recruitment of four initial FS schools • Finalise schools Memorandum of Understanding • Plan and book retreat to develop programme and research protocol
Feb	Theoretical development	• Structured literature review (around elements of a logic model) looking at spaced learning from both a neuroscientific and cognitive psychology perspective
Mar/Apr/May		• Design initial logic model(s) • Draft program development retreat agenda • Capacity building on logic modelling (QUB) • Literature review presentation (QUB) • Discuss 'pure control' • Recruit 12 schools for OS trial
Jun	Feasibility study (FS)	• Group of teachers worked with the evaluation and project teams to develop activities, PowerPoints, training, and lesson plans • Draft training manual
July/Aug/Sep		• Reflect on the quality of the materials, whether or not they are useful and engaging, etc. • Review manual including: literature review, logic model, training manual, and delivery materials
Oct		• FS conducted with 4 schools • Developed bespoke science test based on past GCSE science papers
Nov		• Developed implementation survey
Dec		• Analyse data • Review logic models and materials based on pilot
2016		
Jan	Optimisation study	• Pre-test in 12 OS schools with bespoke science test
Feb		• Training and implementation of programme
Mar		• Training and implementation of programme
Apr		• Post-test measures, bespoke science test an implementation measure
May		• Analyse OS data
Jun	Review and finalise	• Present final data
July		• Revise programme manual (including literature review, logic model, training manual, and delivery materials) • Programme manual graphic design
Aug/Sep		• Finalise research report and submit to the EEF
Oct		• Finalise SMART Spaces manual and submit to the EEF

Findings

Feasibility study

This section summarises the key findings from the FS, which evaluates whether the intervention is likely to be practical and feasible. In essence, it answers the first research question:

> What can we learn about the feasibility of implementing a 'spaced learning' programme in a practical classroom situation?

In order to do this, interviews were conducted with eight teachers and headteachers who had a range of previous experience with spaced learning. Focus groups were also conducted with five children in each of the four participating schools in the FS (20 children in total). The analysis of qualitative data and subsequent programme adaptations were based on emergent themes rather than on any one individual's comments.

Outline of qualitative results from the feasibility study

We received substantial positive feedback from both teachers and pupils about the programme. The qualitative data provided insight into the benefits and limitations of spaced learning, as seen by pupils and teachers. As we progressed towards the OS stage, the interviews and workshops highlighted some specific areas for adaptation to the programme delivery, materials, and training.

- **Programme adaptations**
  - Teachers reported the importance of needing to receive the resources well in advance, as, in numerous instances, materials arrived too close to planned delivery days, or in one instance, the juggling balls arrived a day late.
  - Teachers and headteachers reported that last minute changes-of-plan did not consider the busy school environment. Any school visits or scheduling changes must be planned as far in advance as possible.
  - Juggling: one teacher reported that children who struggled to learn this skill 'gave up' and were frustrated or bored by the third lesson. Teachers suggested other distractor activities such as balloon games, transformer toys, and origami (all used successfully before). Balls were reportedly of poor quality.[5]

ACTION for the OS: Timely delivery of materials and resources are needed to reduce pressure on staff. Consideration should be given to expanding the training relating to juggling to help teachers maintain pupil engagement with both the activity and the lesson.

- **Material adaptations**
  - The length of the PowerPoint (PPT) presentations was inconsistent between topics and subjects, according to both children and teachers. Many teachers felt that the presentations were too long, and it was a difficult to fit them, and juggling, into an hour's lesson.
  - Several teachers suggested that the materials would benefit from more proof-reading, and that some of the science content should be double-checked and updated to ensure it meets the most recent curriculum specifications.
  - There were some problems with visuals on the biology slides, for example, some of the pictures were covered over.

[5] Analysis of the OS would not suggest that using juggling balls has a deleterious influence on effects, so juggling will be still be recommended in the SMART Spaces manual as the space distractor task. However, instruction will be provided in SMART Training to ensure pupil engagement in the task.

- The synchronisation of slides was sometimes off. For example, a missing word would appear on a slide, children would shout the answer, but the correct answer would not appear and the PPT skipped straight to a new gap. Teachers reported having to modify the PPT structure to ensure it ran smoothly.
- 'Each PPT was a different style…varied in length, varied in animations…it needs to be a consistent design', reported one teacher.
- Pupils and teachers reported that the layered style of the slides seemed 'old fashioned'.

Action for the OS: PPTs need to be checked to ensure that *all* variants fit within one hour, including breaks. They also require extra proof reading. The flow and synchronisation of slides needs to be checked, and similar styles between subjects will give a more consistent experience.

- **Training adaptations**
  - Teachers reported that they should have been advised not to have three different teachers running the three different portions of the '24-hour' lessons. When this happened it resulted in organisational problems, made the lessons harder to follow, and the children felt they were rushed and less engaged.
  - Some teachers reported that the set-up meeting between teachers, evaluation team, and project team did not provide sufficient training for this project.[6] Those with prior experience of spaced learning were not comfortable with other variants of the lessons. Further modelling and practice of the different programme variants could have been provided.
  - Pupils noted that different teachers showed different levels of comfort with delivery. This may be reflective of the fact that not all teachers were directly trained.
  - Slides in training were different from the slides used during delivery. Teachers expressed a need for training using the specific slides they would be using. One teacher felt that there was an assumption that teachers would know the 'story' of the presentation, and how things were connected, when this was not readily apparent.
  - One teacher reported not having had training specific to the project. This made the 24-hour space version very difficult to teach. They had only been trained on the ten-minute space version.
  - Numerous teachers suggested that getting a chance to practice spaced learning, with feedback, would be very useful—perhaps more examples of 'good' spaced learning practice versus 'bad' practice. Teachers felt there should be a 'standardised model' of the spaced learning variants to refer to.
  - Teachers who had to train other members of staff within their department felt this was not adequate.
  - Children reported that some teachers tried to add in too much extra information to the slides. This could be addressed in training.

Action for the OS: All teachers delivering the programme need to be trained. Training needs to specify that one teacher will deliver all lessons within a variant, even if the topic is not their specialist subject. Teachers need to be given more confidence in delivering a 'good' spaced learning lesson, and need to come away knowing what a standardised spaced lesson looks like. Training should use the exact materials the teachers will be delivering. All these concerns are now addressed in the SMART Training session and SMART Spaces manual.

- **Feedback on specific variants**

[6] Teachers in the FS were only given information about the programme and delivery at the programme development retreat. SMART Training was developed and implemented at the OS stage.

- **Variant 1 (ten-minute spaces).** Teachers felt most comfortable delivering this as many had prior experience of spaced learning in this style. The children who received these lessons reported that their teachers seemed confident and enthusiastic—more so than children receiving other variants.
- **Variant 2 (24-hour spaces)**. Children and teachers felt this variant was more rushed and difficult to deliver as some schools attempted to use three different teachers for the three different subjects. The concept of 24-hour spaces between lessons seems to have been less noticeable to participants than the challenge of squeezing three teachers into one lesson. The lack of training, and lack of prior experience of this variant, made teachers less confident.
- **Variant 3 (ten-minute and 24-hour combination)**. Children reported enjoying the juggling breaks and feeling engaged, but teachers reported that this variant was too long and too difficult to fit into one hour. Ensuring that only one teacher delivers this lesson should help with timing issues, but content editing for all variants seems necessary due to the frequent reports of difficulty fitting into one hour. One teacher reported finding this variant tiring to deliver.

Action for the OS: It was judged that although there were difficulties in delivering specific variants these could be overcome by the adaptations to the programme, materials, and training as described above. Furthermore, it was deemed necessary to optimise the programme through a preliminary check of the efficacy of the different variants.

Performance of measures

One hundred and sixty-nine pupils completed the bespoke GCSE science assessment at post-test in the FS. Reliability tests showed a Cronbach's alpha score of 0.85. Furthermore, total scores on the test were found to be normally distributed thus indicating no floor or ceiling effects of the test. In addition, no pupil scored 0 or the maximum of 57.

Optimisation study

Participants

Twelve schools participated in the optimisation study phase. Table 7 shows five key characteristics for each school: (1) whether urban or rural, (2) the size of the school, (3) the proportion of pupils in receipt of free school meals ('FSM pupils'—an indicator of level of deprivation), (4) the type of school, and (5) its Ofsted rating.

Table 7: The characteristics of the 12 schools involved in optimisation study

School	Variant Type	Urban/ Rural	Size of School (n)	FSM Pupils (%)	Type of Establishment	Ofsted
A	No spaces	Urban	1755	15	Academy Converter	Good
B	10-min	Rural	1522	3	Academy Converter	Outstanding
C	10-min & 24-hr	Urban	725	15	Community School	Outstanding
D	No spaces/materials	Urban	525	21	Academy Converter	Good
E	10-min & 24-hr	Urban	1650	15	Academy Converter	Good
F	10-min & 24-hr	Urban	980	21	Academy Converter	Good
G	No spaces	Urban	1244	15	Academy Converter	Good
H	24-hr	Urban	1600	10	Community School	Good
I	10-min	Rural	600	16	Academy Converter	Good
J	10-min	Urban	600	12	University Technical College	Good
K	24-hr	Urban	1365	35	Academy Sponsor Led	Good
L	No spaces/materials	Urban	160	0	Other Independent School	N/A

Evidence to support theory of change

Comparison of variant effects

To get an indication of effects of the programme variants over the different groups, we compared the effect sizes (ES = Hedge's g using mean gain scores, pre and post SD's, and paired t-tests) of the different conditions and compared them to the control (Table 8).[7]

It can be seen that Variant 3 (the combined ten-minute and 24-hour variant) showed a pattern of positive effects on all scores and against both controls, with one significant difference showing greater positive change between Variant 3 (total score) and the no spaces/materials control (ES = 0.19). There was a more modest effect of the Variant 1 (ten-minute version) across the scores and controls with no significant effect on any of the comparisons. Variant 2 (24-hour version) produced some low and some negative effects across all comparisons with one significant negative effect between the Variant 2 long answer score and slide-only control (ES = -0.37). Therefore, it can be seen that Variant 3 shows the best and most consistent evidence of promise against controls at this stage of programme development.

[7] Appendix 8 provides a table showing pre, post, and gain score summary (across the five groups). This table also includes gain score removing school L which was reallocated to the 'no spaces/materials' control group. It suggests that school L had better gains than the other school in this control group. Therefore, if anything, school L dampened the effect size difference between control and the three variant types.

Table 8: Effect sizes for the three variants on total, short, and long answer scores compared against the two control groups

Variant type	Control group type	Test score type	Intervention change (pre s.d., post s.d., N, paired t)	Control change (pre s.d., post s.d., N, paired t)	Effect size = *g* (95% confidence intervals)
10-min	Slides only	Total	2.89 (9.09, 9.82, 100, -5.49)	2.62 (8.54, 8.89, 68, -4.39)	0.03 (-0.14, 0.20)
10-min	Slides only	Short	2.26 (6.25, 6.58, 100, -5.47)	1.62 (6.05, 6.60, 68, -2.57)	0.10 (-0.13, 0.33)
10-min	Slides only	Long	0.63 (4.07, 4.41, 100, -2.47)	1.00 (3.56, 3.72, 68, -2.43)	-0.09 (-0.34, 0.16)
10-min	Nothing	Total	2.89 (9.09, 9.82, 100, -5.49)	1.82 (7.16, 9.79, 50, -2.48)	0.12 (-0.07, 0.30)
10-min	Nothing	Short	2.26 (6.25, 6.58, 100, -5.47)	1.42 (5.41, 6.94 50, -2.80)	0.13 (-0.08, 0.34)
10-min	Nothing	Long	0.63 (4.07, 4.41, 100, -2.47)	0.40 (2.88, 3.90 50, -0.90)	0.06 (-0.21, 0.33)
24-hr	Slides only	Total	1.54 (12.81, 14.63, 69 -2.36)	2.62 (8.54, 8.89, 68, -4.39)	-0.09 (-0.25, 0.07)
24-hr	Slides only	Short	2.03 (9.23, 10.81 69, -3.66)	1.62 (6.05, 6.60, 68, -2.57)	0.05 (-0.17, 0.27
24-hr	Slides only	Long	-0.49 (4.13, 4.62, 69, 1.65)	1.00 (3.56, 3.72, 68, -2.43)	**-0.37*** (-0.63, -0.11)
24-hr	Nothing	Total	1.54 (12.81, 14.63, 69 -2.36)	1.82 (7.16, 9.79, 50, -2.48)	-0.02 (-0.20, 0.14)
24-hr	Nothing	Short	2.03 (9.23, 10.81 69, -3.66)	1.42 (5.41, 6.94 50, -2.80)	0.07 (-0.11, 0.25)
24-hr	Nothing	Long	-0.49 (4.13, 4.62, 69, 1.65)	0.40 (2.88, 3.90 50, -0.90)	-0.22 (-0.50, 0.06)
10-min & 24-hr	Slides only	Total	3.76 (10.59, 12.22, 76 -6.55)	2.62 (8.54, 8.89, 68, -4.39)	0.11 (-0.05, 0.28)
10-min & 24-hr	Slides only	Short	2.53 (8.10, 8.70, 76, -5.65)	1.62 (6.05, 6.60, 68, -2.57)	0.12 (-0.11, 0.35)
10-min & 24-hr	Slides only	Long	1.24 (3.65, 4.30, 76, -3.88)	1.00 (3.56, 3.72, 68, -2.43)	0.06 (-0.21, 0.33)
10-min & 24-hr	Nothing	Total	3.76 (10.59, 12.22, 76, -6.55)	1.82 (7.16, 9.79, 50, -2.48)	**0.19*** (0.01, 0.36)
10-min & 24-hr	Nothing	Short	2.53 (8.10, 8.70, 76, -5.65)	1.42 (5.41, 6.94, 50, -2.80)	0.15 (-0.03, 0.33)
10-min & 24-hr	Nothing	Long	1.24 (3.65, 4.30, 76, -3.88)	0.40 (2.88, 3.90, 50, -0.90)	0.22 (-.07, 0.51)

*Significant < 0.05

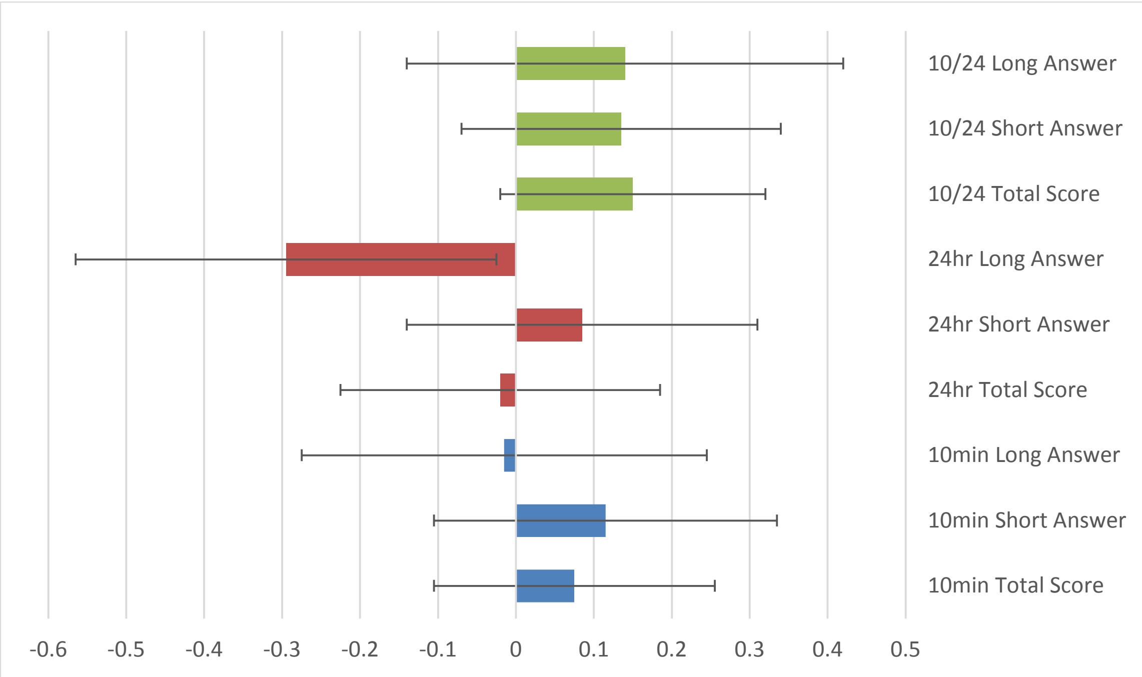
To further demonstrate the pattern of effects, Figure 2 shows an error bar chart (with 95% confidence limits) of the mean effect size, comparing the mean variant effect size across both controls.

For example:

> [(ten-minute total versus 'slides-only' control, ES = 0.03) + (ten-minute total versus 'no spaces/materials' control, ES = 0.12)] divided by 2 = 0 .075.

Again, Figure 2 shows a clear pattern of higher effect sizes for Variant 3 against the other two versions. In fact, it is particularly beneficial on the long answer questions over the other two conditions. It is worth noting the particularly poor performance of the ten-minute and 24-hour versions on long answer questions, with the 24-hour version showing a large significant negative effect on these questions.

Figure 2: Error bar chart (with confidence intervals) showing a comparison of mean variant effect size compared to both controls, for all scores (total, short answer, and long answer)



Implementation Factors

The next step was to examine the relationship between contextual factors and effects of the approach.

Engagement

One of the important contextual factors reported by the project team was the engagement of teachers and pupils in the programme. Therefore, a spaced learning engagement measure was developed and administered to the pupils at post-test reporting on their teachers' and their own engagement in the programme. This measure was found to be reliable (Cronbach's Alpha = 0.91) and was regressed onto outcome change (Table 9). Two hundred and twenty-four pupils received one of the three variants of the intervention, completed pre- and post-tests, and the engagement questionnaire. On analysis of the relationship between their outcomes and engagement it can be seen that this engagement score was a significant implementation factor, with higher engagement score 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).

Table 9: Regression of independent variables pre-test and engagement onto post-test outcome scores

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-0.97	2.04		-0.47	0.64
	Pre-test total score (max=57)	0.99	0.03	0.40	30.54	<.01
	Engagement score	0.11	0.05	0.06	2.13	0.03
Dependent Variable: post-test total score (N = 224) across all variant types (max score = 57).						

Inter-study activity

Another frequently occurring question concerned what activity was best to use in the ten-minute 'spaces'—was it juggling or some other activity? Some schools did use activities other than juggling, but only a low number (juggling n = 214, other activity n = 18). An independent samples t-test showed that those who used juggling had significantly higher outcome change scores (juggling mean = 3.09 and s.d. = 5.03) compared to using another activity (mean = -0.83 and s.d. = 5.14) (t(df) = 3.17 (230), p = <0.01). However, this result must be viewed with caution given the small number who did not use juggling and the fact that this could be a proxy for engagement as all teachers were instructed to use juggling in the ten-minute spaces. Furthermore, pupils were reporting using an activity other than juggling across a number of schools, although it seemed to be one school that had the majority of pupils reporting other activity (N = 8). There may have been a mix of use of other activities and reporting errors. This result must therefore be treated with caution.

Retention

Previous literature has suggested that the length of spaces has an influence on the period of retention of information. The final factor investigated was the influence of retention on programme effects. There was some variation in the time between the programme variant ending and the post-test occurring (due to Easter holidays). Of the 313 pupils who completed pre- and post-tests and received a programme variant (this includes the 'slides-only' control group, but not the 'business as usual' control group), there was a mean interval of 18.19 days (s.d. = 12.44) between the end of the programme and the test date with a minimum gap of seven days and maximum gap of 48 days reported. This variation was exploited to explore the influence of retention interval. To do this, a regression was conducted which showed no significant effect of retention interval on the post-test scores (see Table 10).

Table 10: Regression of retention interval and pre-test score onto post-test score outcome variable

Model		Unstandardized coefficients		Standardized coefficients	T	Sig.
		B	Std. error	Beta		
1	(Constant)	3.520	0.969		3.632	<0.001
	Pre-test - total score (max = 57)	0.977	0.028	0.893	34.440	<0.001
	Time lapsed between programme end and post test	-0.012	0.025	-0.012	-0.457	0.648
Dependent Variable: post-test - total score (N = 313) across all variant types (max score = 57).						

Outline of qualitative results from the optimisation study

Interviews were conducted with teachers who delivered each of the spaced learning variants (not including control conditions). The qualitative data of teachers' perspectives from this optimisation study shows that the actions determined by the feasibility study have been addressed successfully.

Training and support from the delivery team

- Teachers reported that the training session was very enjoyable. The small size of training groups was appreciated as it facilitated extensive personal feedback and a chance to refine lesson delivery.
- One teacher suggested that the training could have focused more on how to explain to the pupils the 'science behind' spaced learning to encourage engagement with the project.
- The face-to-face nature of the training session was perceived as very appropriate, as training for this style of teaching, 'definitely does need to be in person'. Teachers reported receiving useful advice, for example, to restrict ad-libbing and stick to materials for the purposes of the research.
- The successful training meant that most teachers did not feel they needed much further support to deliver the intervention.

Materials

- Teachers reported that the PowerPoint slides were very high quality, and although they didn't always fit exactly with their exam board curriculum, they felt comfortable using the slides in the lessons. Others felt the content of the slides was not consistently engaging.
- A few spelling errors and American spellings were noticed.
- One teacher suggested that today's pupils are used to slightly more 'high tech' lesson materials, and that something more engaging and interactive than PowerPoint slides could be used.
- Versions of the slides for the visually impaired should be available on request.

Fidelity of delivery

- Teachers generally reported that they stuck exactly to the provided materials. No problems were reported relating to running the three days, aside from some minor timetable changes.
- Some adaptations had to be made to PPT slides, such as spelling, definitions that did not fit with the children's knowledge of the curriculum, and some colour/visual changes.
- One teacher reported switching to 'paper balls' to stop children bouncing the juggling balls around the classroom, and problems with balls disappearing at the end of class.

Pupil engagement and wider impacts

- Many teachers reported that they felt pupils in general engaged well with the lessons.
- Low ability students were considered more likely to 'get more out of it'.
- Higher ability students—A*, A, and top B—were considered more likely to get bored due to repetition, or to feel like the material wasn't sufficient to improve their grade.
- One teacher reported that his higher set actually engaged better with spaced learning, but reported that they 'didn't feel like they had learned anything'.
- Numerous teachers have gone on to continue running spaced learning with other classes and subjects after the project. One suggested that they will keep using spaced learning, but not with juggling—maybe 'mindful colouring'.
- They reported being able to use the layout of the slides to suit their syllabus, and that it was great to be able to hold on to them and use them for further spaced learning in the school.

Feedback on specific variants

Variant 1 (ten-minute spaces). One teacher reported no major issues with implementing this lesson style and that it ran smoothly. Another, however, reported that their lower ability students found the pace too difficult. Class control was an issue for one teacher due to the frequent bouts of physical activity in the class.

Variant 2 (24-hour spaces). The first session, which is the most content-heavy in the variant, was difficult to run and was longer than later sessions. One teacher reported that students really 'latched on' to the lessons towards the end of the program, despite engagement difficulties early on.

Variant 3 (ten-minute and 24-hour). No time challenges were reported for the first session in this variant, although it includes the same depth of content as the first session in Variant 2. It may be the case that the juggling 'spaces' help to address the difficulty of having the three longer blocks of content from each subject in the one session. One teacher, however, reported that juggling throughout the lesson was difficult for some students, and that they began to feel shy about struggling with it.

Conclusion

Key conclusions	
1.	The spaced learning principle is based on relatively strong evidence and this pilot suggests that **SMART Spaces** has evidence of promise.
2.	This pilot study demonstrated that **SMART Spaces** can feasibly be delivered in English schools. Both teachers and pupils appeared to enjoy and engage with the programme.
3.	Teachers generally reported that they delivered **SMART Spaces** lessons as prescribed and did not make major alterations.
4.	The small randomised controlled trial provided some preliminary evidence that the most promising approach to spaced learning combines both ten-minute and 24-hour spaces. However this was a small study and a larger trial is needed to fully understand the impact of the programme.
5.	**SMART Spaces** is ready for a larger RCT to evaluate its impact on GCSE attainment.

Formative findings

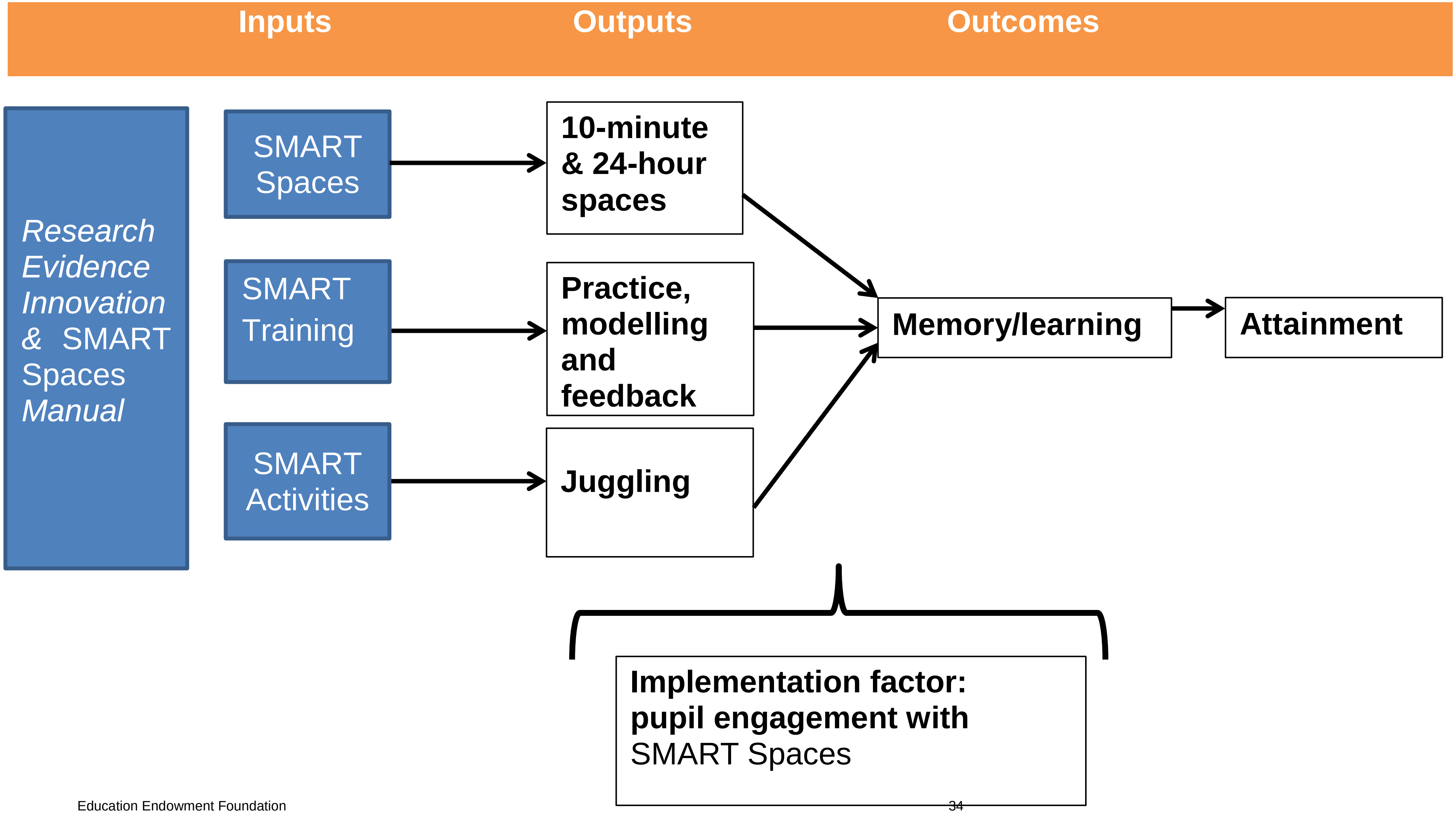
As indicated earlier, the programme has had two iterations of change. The first changes came after the formative findings from the FS, which saw adaptations to the materials, delivery, and training. Similarly, after the OS the evaluation and project teams concluded that the ten-minute and 24-hour combined version of the programme shows the best evidence of promise for improving science attainment. Furthermore, this version is underpinned by evidence from both the neuroscientific and cognitive experimental literature. Therefore it has been chosen as the core model for the SMART Spaces programme. The final version of the spaced learning programme in presented in Figure 3.

It is important to emphasise the training element of the programme (SMART Training), particularly in the light of a recent review of the effectiveness of secondary science education programmes (Cheung *et al.*, 2016). This review highlights that effective secondary science programmes require a strong focus on professional development and support for teaching, rather than simple material-focused innovations.

Key recommendations with regard to training are:

- All teachers that deliver the programme need to be trained by an experienced project team. Even teachers who have some experience of spaced learning may not be familiar with the specific delivery model that SMART Spaces requires.
- SMART Training should include a significant practice element where the teachers are provided with the opportunity to deliver a spaced learning lesson and receive feedback on their performance.
- One teacher should provide all lessons within the programme, for two reasons: (1) to build their confidence in the materials, and (2) to ensure that the programme is completed in the allotted time with the minimum disruption to timetabling and teacher availability.
- The juggling activity should be supported by some training guidance on how to run this session successfully.

Figure 3: SMART Spaces logic model



Interpretation

Previous literature has shown the benefits of both short spaces (around ten minutes) and medium-term spaces (24 hours plus) on the retention of information. Furthermore, some educational practice literature shows the benefits of both these kinds of spaces in a real educational setting. However, the emerging picture from this research would suggest that there is a benefit to combining both short ten-minute and medium-term 24-hour spaces for improving attainment outcomes. Coupled with the fact that these lessons can be feasibly delivered in classroom settings makes it a potentially powerful way to conduct revision for GCSE science exams.

As mentioned in the rationale, the aim of this study was to consider how the theory relating to spaced learning might translate into a practical model—an optimal spacing pattern—for improving attainment outcomes. This has resulted in the ten-minute and 24-hour model underpinning SMART Spaces. The theory of change, in terms of the neurological and cognitive changes that occurred as a result of the spaces, was not explored.

However, one interesting finding with regard to the 24-hour variant producing a detrimental effect may lead to one hypothesis on the theory of change: the 24-hour variant did not provide any spaces within each daily session of learning and thus provided substantial amounts of information in a massed way in each session over the three days. Teachers who provided this version commented on the fact that this was difficult and tiring for the pupils. Looking at the neuroscientific literature, it has been observed that short-term spacing may produce effects on memory through neurochemical refreshment (for example, refreshment of proteins such as CREB) and/or the benefits of switching to novel tasks. It may be that the mechanism of change was not in operation in the 24-hour version and yet is important in producing the effects in spaced learning practice. It may be hypothesised that the short spaces provide some neurobiological advantage that can be built on by the 24-hour spaces that were also included in the ten-minute and 24-hour combined variant. The hypothesised theory of change, therefore, is that proximal changes produced by short spaces are required to build on the distal benefits of longer spaces in this way of teaching. In the cognitive psychology literature, 24-hour spaces have been shown to be an effective spacing strategy for retention. The learning experience, however, may be more difficult as the long session of intake of material without breaks may cause fatigue. It would seem that when the ten-minute breaks are combined with these 24-hour spaces, the fatigue is reduced and the benefits of 24-hour spaces may be revealed. This is speculation at this point: other explanations of the added value of the ten-minute space may be related to the reduction in interference between the different subject areas taught. Further neuroscientific and cognitive experimental study would be required to investigate the theories of change operating in SMART Spaces.

Another observed pattern was that the variants performed better against the pure control group than the 'slides-only' group. This suggests that there is some intrinsic benefit in the way the content is presented in a high intensity manner. Therefore, it is important that the slides are of high quality and updated regularly based on the current curriculum.

Finally, it is worth noting that both the 24-hour and the ten-minute and 24-hour combined version of spaced learning included elements similar to interleaving as they used slightly different content in the one session. There are many further nuances to interleaving research (Taylor and Rohrer, 2010) that were not explored in this study that would need further investigation.

There are several limitations to this research. First, the sample size is relatively small for a trial, and not adequately powered to detect expected effects, and 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 particular variant being delivered. In addition, the measures used, although found to be reliable and demonstrating validity, do not provide actual GCSE science scores. Furthermore, this programme was implemented as a revision programme, so it needs additional testing for its contribution to initial learning. Finally, this was not a well-controlled RCT, and increased rigour in this respect will clarify the effectiveness of SMART Spaces.

Readiness for trial

The evaluation team suggest that SMART Spaces is ready for an efficacy trial. The process of design and feasibility has provided substantial learning on the feasibility of the programme's implementation, as well as preliminary evidence of its promise for improving science attainment. Furthermore, the measure to assess its effect performed well (although actual GCSE scores may be a better outcome measure for a larger trial). In addition, there seems to be good reliability and validity to the engagement measure—this would be a useful tool for monitoring and assessing implementation in a larger trial.

A contribution to the scalability and replicability of SMART Spaces is the development of a manual. The manual will be informed by the findings in this study and help ensure fidelity of delivery going forward. Some suggestions for the design and research questions in a larger trial are discussed in the next section on future research.

Another necessary consideration for going towards a bigger trial and potentially greater roll-out of the programme is cost. The cost per pupil is £10. This cost is a one off cost to get the teacher trained and provide the school with the necessary materials to run the programme for three years. In theory, this cost is for the first cohort of pupils, but will also cover subsequent cohorts of pupils being taught by the same teacher. The cost per pupil is therefore decreasing with each cohort that goes through, however changes in staff, and to the GCSE curriculum, may require refresher packs and training sessions for schools.

Future research and publications

The evaluation team suggest an efficacy trial at the next stage with further exploration of feasibility using both qualitative and quantitative measures (particularly the engagement questionnaire included in this study, which can help monitor this significantly related implementation factor). A key unanswered question at this point is the cumulative effect of SMART Spaces if used throughout the school year. This question could be answered by having a three-arm trial of SMART Spaces with the programme occurring at the beginning and end of the final term of the GCSE course compared to twice at the end of the course just before exams. This design would provide the opportunity to compare the use of ten-minute, 24-hour, and a whole-term space against the current ten-minute and 24-hour model of SMART Spaces. It is also recommended that the current bespoke measure is used as a pre-test, and actual GCSE scores are used as the post-test outcome measure. Some suggested research questions for future investigations include:

1. What are the effects of SMART Spaces on GCSE science attainment?
2. Does implementing SMART Spaces at several time-points (at the beginning and end of final term) have an additional beneficial effect on GCSE science attainment?
3. Can SMART Spaces be delivered with fidelity (and engagement) in a scaled up version of the programme?

The evaluation team are currently co-producing a SMART Spaces training manual and SMART Training programme that will be a useable guide for trainers and teachers in the delivery of the SMART Spaces programme. They are also intending to publish some of the results in this report in an international high quality academic journal with potential further papers in science education practice journals.

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.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/23521365

Bird, S. (2011) 'Effects of distributed practice on the acquisition of second language English syntax', *Applied Psycholinguistics*, 32, pp. 435–452.

Bonell, C., Fletcher, A., Morton, M., Lorenc, T. and Moore, L. (2012) 'Realist randomised controlled trials: a new approach to evaluating complex public health interventions', *Social science & medicine*, 75 (12), pp. 2,299–306.

Carpenter, S. K. *et al.* (2012) 'Using Spacing to Enhance Diverse Forms of Learning: Review of Recent Research and Implications for Instruction', *Educational Psychology Review*, 24, pp. 369–378.

Carpenter, S. K., Pashler, H. and Cepeda, N. J. (2009) 'Using Tests to Enhance 8th Grade Students Retention of U.S. History Facts', *Applied Cognitive Psychology*, 23 (August 2008). Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-37349131250&partnerID=40&md5=b2db461a1b1b0e6fe1ad3312b05a0329

Cepeda, N. J. *et al*. (2009) 'Optimizing distributed practice theoretical analysis and practical implications', *Experimental Psychology*, 56 (4), pp. 236–246. Available at: http://dx.doi.org/10.1027/1618-3169.56.4.236

Cepeda, N. J. *et al.* (2006) 'Distributed practice in verbal recall tasks: A review and quantitative synthesis', *Psychological bulletin*, 132 (3), pp. 354–380.

Cheung, A., Slavin, R. E., Lake, C. and Kim, E. (2016) 'Effective secondary science programs: A best-evidence synthesis', in *Annual meeting of the Society for Research on Educational Effectiveness*, Washington, DC.

Commins, S. *et al.* (2003) 'Massed but not spaced training impairs spatial memory', *Behavioural Brain Research*, 139, pp. 215–223.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. and Petticrew, M. (2008) 'Developing and evaluating complex interventions: the new Medical Research Council guidance', *BMJ*, 337, p. a1655.

Dempster, F. N. (1988) 'The spacing effect: A case study in the failure to apply the results of psychological research', *American Psychologist*, 43 (8), p. 627.

Dunlosky, J. *et al*. (2013) 'Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology', *Psychological Science in the Public Interest*, 14, pp. 4–58. Available at: http://psi.sagepub.com/lookup/doi/10.1177/1529100612453266

Fields, R. D. (2009) 'Making memories stick', *Scientific American*, 120 (February), p. 5.

Gerbier, E. and Toppino, T. C. (2015) 'The effect of distributed practice: Neuroscience, cognition, and education', *Trends in Neuroscience and Education*, 4 (3), pp. 49–59. Available at: http://dx.doi.org/10.1016/j.tine.2015.01.001

Gui Xue, Leilei Mei, Chuansheng Chen, Zhong-Lin Lu, Russell Poldrack, and Qi Dong (2012) 'Spaced Learning Enhances Subsequent Recognition Memory by Reducing Neural Repetition Suppression', *Journal of Cognitive Neuroscience*, 23 (7), pp. 1,624–633.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M. and Lamb, S.E. (2014) ‘Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide’, *BMJ*, 348, p. g1,687.

Huff, M. J. & Bodner, G. E. (2014) ‘All varieties of encoding variability are not created equal: Separating variable processing from variable tasks’, *Journal of Memory and Language*, 73, pp. 43–58. Available at: http://dx.doi.org/10.1016/j.jml.2014.02.004

Jackson, C. E., Maruff, P. T. and Snyder, P. J. (2013) ‘Massed versus spaced visuospatial memory in cognitively healthy young and older adults’, *Alzheimer’s and Dementia*, 9 (1), pp. S32–S38. Available at: http://dx.doi.org/10.1016/j.jalz.2012.04.005

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’, Journal of Consumer Research, 30 (1), pp. 138–149.

Jamal, F., Fletcher, A., Shackleton, N., Elbourne, D., Viner, R. and Bonell, C. (2015) ‘The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example’, *Trials*, 16 (1), p. 1.

Kapler, I. V, Weston, T. and Wiseheart, M. (2015) ‘Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning’, *Learning and Instruction*, 36, pp. 38–45. Available at: http://dx.doi.org/10.1016/j.learninstruc.2014.11.001

Kelley, P. (2009) ‘Spaced Learning: A harbinger of a paradigm shift in education’, *Education Today*, 59, pp. 3–6.

Kelley, P. and Whatson, T. (2013) ‘Making long-term memories in minutes: a spaced learning pattern from memory research in education’, *Frontiers in human neuroscience*, 7 (Sep.), p. 589. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3782739&tool=pmcentrez&rendertype=abstract

Kerfoot, B. P. (2010) ‘Adaptive Spaced Education Improves Learning Efficiency: A Randomized Controlled Trial’, *Journal of Urology*, 183 (2), pp. 678–681. Available at: http://dx.doi.org/10.1016/j.juro.2009.10.005

Kogan, J. H. *et al*. (1997) ‘Spaced training induces normal long-term memory in CREB mutant mice’, *Current biology*, 7, pp. 1–11.

Kornmeier, J. and Sosic-Vasic, Z. (2012) ‘Parallels between spacing effects during behavioral and cellular learning’, *Frontiers in Human Neuroscience*, 6 (July), pp. 1–5.

Kramar, E. A. *et al*. (2012) ‘Synaptic evidence for the efficacy of spaced learning’, *Proceedings of the National Academy of Sciences*, 109 (13), pp. 5,121–126.

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’, *Learning & Memory*, 5, pp. 246–256.

Menzel, R. *et al.* (2001) ‘Massed and Spaced Learning in Honeybees: The Role of CS, US, the Intertrial Interval, and the Test Interval’, *Learning & Memory*, 8, pp. 198–208.

Miles, S. W. (2014) ‘Spaced vs. massed distribution instruction for L2 grammar learning’, *System*, 42, pp. 412–428. Available at: http://dx.doi.org/10.1016/j.system.2014.01.014

Mollison, M. V and Curran, T. (2011) ‘Investigating the Spacing Effect Using EEG’.

Moss, V. D. (1995) ‘The efficacy of massed versus distributed practice as a function of desired learning outcomes and grade level of the student’, *Doctoral dissertation, ProQuest Information & Learning*.

Rohrer, D. (2012) ‘Interleaving Helps Students Distinguish among Similar Concepts’, *Educational Psychology Review*, 24, pp. 355–367.

Toppino, T. C. and DiGeorge, W. (1984) ‘The spacing effect in free recall emerges with development’, *Memory & cognition*, 12 (2), pp. 118–122.

Van Strien, J. W. *et al*. (2007) ‘Electrophysiological correlates of word repetition spacing: ERP and induced band power old/new effects with massed and spaced repetitions’, *International Journal of Psychophysiology*, 66, pp. 205–214.

Zhang, Y. *et al.* (2011) ‘Computational design of enhanced learning protocols’, *Nature Neuroscience*, 15 (2), pp. 294–297. Available at: http://dx.doi.org/10.1038/nn.2990

Appendix 1: Summary of earlier key studies

Study type	Authors (year)	Input	Output	Outcome
Cognitive psychology experiment (pre-schoolers and 1st graders)	Toppino & DiGeorge (1984)	Access to nursery and 1st grade children. One to one testing with researcher and child. Testing of 20 children individually clearly less cost effective than classroom based study.	Free recall experiments comparing massed and spaced repetitions of words with nursery and first grade children, of 4 different word lists (each list contained 3 words which were repeated in a massed format, and 3 repeated in a spaced format).	First grade children showed a clear spacing effect, whereas pre-school children did not. It is possible that attention is consciously diverted from repeated stimuli in massed repetitions. *This is where ERP research would be very useful.*
Review paper	Lee & Genovese (1988)	*Review paper.*	47 articles on effect of distributed practice on acquisition and retention of motor skills.	Distributed practice improved both acquisition and retention of motor skills.
Review paper	Moss (1996)	*Review paper.*	Review of 120 articles on distributed practice effect, across age of participant and material type (verbal information, intellectual skill, motor learning)	Longer ISIs (inter study intervals) helped verbal information (e.g. spelling) and motor learning (e.g. mirror tracing) in over 80% of studies. Only one third of studies showed that spaced practice helped intellectual skill (e.g. math computation) and half of them showed no effect.
Neuroscience – Animals	Kogan et al. (1997)	Animal research lab facilities, raising of genetically mutated mice.	Comparison of spaced and massed learning with mice who lacked the CREB protein (a protein important for creation of long term memories).	Spaced training with ISIs of 60 minutes allowed typical memory performance of mice who otherwise were unable to form long term memories for spatial, contextual and social information.
Neuroscience - Animals	Mauelshagen, Sherff & Carew (1998)	Animal research lab facilities – Aplysia (marine molluscs).	Comparison of spaced and massed serotonin exposure with aplysia. Recorded intracellular firing of synapses using	Long term synaptic facilitation (electrical activity recorded up to 24 hours later) was significantly increased when serotonin was applied in 5 x 5

			electrodes.	minute exposures with 15 min intervals, compared with massed exposure of 25 min constant application.
Review paper	Donovan & Radosevich (1999)	*Review paper.*	63 articles reviewed for wide range of tasks. Examined effects of methodological rigour (3 point scale), mental requirements (low or high 'mental or cognitive skills'), overall complexity, ISI (<1 min, 10min - 1 hr, >1day) and retention interval (< or > 1 day).	The largest effects were found in low rigour studies with low task complexity (e.g. typing). The retention interval did not show any effects. Task domain moderated the distributed practice effect. They only examined one interaction, ISI and task domain; Increasing ISI led to larger effect sizes in verbal tasks like verbal discrimination and foreign language. This formed a U-shaped curve, where too long of an ISI led to a smaller effect. Skill based tasks like typing or gymnastics showed smaller effect sizes when intervals were increased.
Neuroscience - Animals	Menzel, Manz, Menzel & Greggers (2001)	Animal research lab facilities – honeybees.	Massed vs spaced learning - Conditioning of honeybees to show proboscis response for feeding, when different CS (odours etc.), US (strengths of sucrose solution) and inter-trial intervals (ITI) are used.	Spaced conditioning led to better acquisition and retention rates, especially for long intervals, i.e. late long term memory, assessed at 4 days (not only higher rates, but more stable). A consolidation process takes place when spaced conditioning is used (they found it could be inhibited by blocking protein synthesis during acquisition).
Review paper.	Janiszewski et al. (2003)	*Review paper.*	97 articles on distributed practice were reviewed.	This study found that verbal vs pictorial stimuli, novel vs familiar stimuli, unimodal vs bimodal stimulus presentation, structural vs semantic cue relationships and isolated vs context

				embedded stimuli did not influence effect size. Five factors influenced effect size magnitude: lag (longer ISI = larger effect), meaningfulness of stimulus = larger effect size than non-meaningful, complex stimulus (i.e. semantically complex) led to larger effect size than structurally complex or simple stimuli, intentional learning larger effect than incidental learning, complex intervening material (i.e. semantically complex) led to larger effect size than structurally complex or simple intervening material.
Computational/ mathematical modelling	Raajmakers (2003)	Programming of computer model of memory.	Tested the SAM (Search of Associated Memory) model	Claimed that spacing effects are due to additional contextual information stored when the material is presented at a second time point (encoding variability). This increases the chance of the context of the testing context being similar to the study context.
Neuroscience – Animals	Commins, Cunningham, Harvey & Walsh (2003)	Animal research lab, rats.	Massed (16 trials) vs spaced (4 trials per day for 4 days). Inter trial intervals of 10 seconds. Training for a water maze and object displacement task.	Spatial memory was impaired in massed relative to spaced trained group. Worse at water maze and worse at detecting object displacement. *Perhaps more of a lag effect, as training was still multiple trials with 10s intervals in massed condition.*
Review paper.	Cepeda et al. (2006)	*Review of 317 experiments in 184 articles. All verbal memory, all test of recall.*	Analysis of spacing vs massed. Analysis of lag effect (comparison of different ISIs).	Only 12 of 217 studies failed to show a benefit of spacing over massed study. Lag effect was nonmonotonic - Increasing ISI increases recall, but

				too long an ISI for a given retention period reduced recall. ISI should increase as retention interval increases to optimise recall. When retention interval was less than 1 minute, optimal ISI was also less than 1 minute. When retention interval was 6 months, at least a 1 month ISI was needed to be optimal. For education, the practical implication is that at least 1 day would be needed for ISI to be optimal for a retention period of months.
Neuroscience - Humans	Van Strien et al. (2007)	EEG lab.	EEG experiment with 22 Dutch adults. Continuous recognition task (old vs new) followed by unexpected free recall task.	Larger N400 (index of memory search) for massed than spaced repetitions. LPC (late positive complex - matching of template) larger massed than spaced repetitions.
Spaced practice – 8th grade students.	Carpenter et al. (2009)	Researcher administered learning sessions during class time.	Children given a history test and then a review of answers/facts. Review of facts was repeated either 1 week or 16 weeks later. Retention test 9 months later.	Children given a review with ISI of 16 weeks recalled more than children given the review with ISI of 1 week. *Note that recall was still very poor for both groups (12% vs 8%) due to large retention period.* Again we see that long retention interval benefits from longer ISI.
Lab based experiment – complex task (language).	Cepeda et al. (2009)	Lab based learning experiments over period of up to 12 months with 215 undergraduates.	Experiment 1 – Learning of Swahili-English word pairs. ISI of 5 min to 14 days, retention interval 10 days. Experiment 2 – Learning of facts and objects, ISI of 20min to 6 months, retention interval 6 months.	Nonmonotonic effect of ISI. Increases in gap led to sharp improvement in retention, then decline. 28 day gap was optimal for 6 month retention interval.
Neuroscience - Animals	Fields (2009)	Research lab – animals.	Stimulation of synapses taken	The synaptic connection, i.e. how it

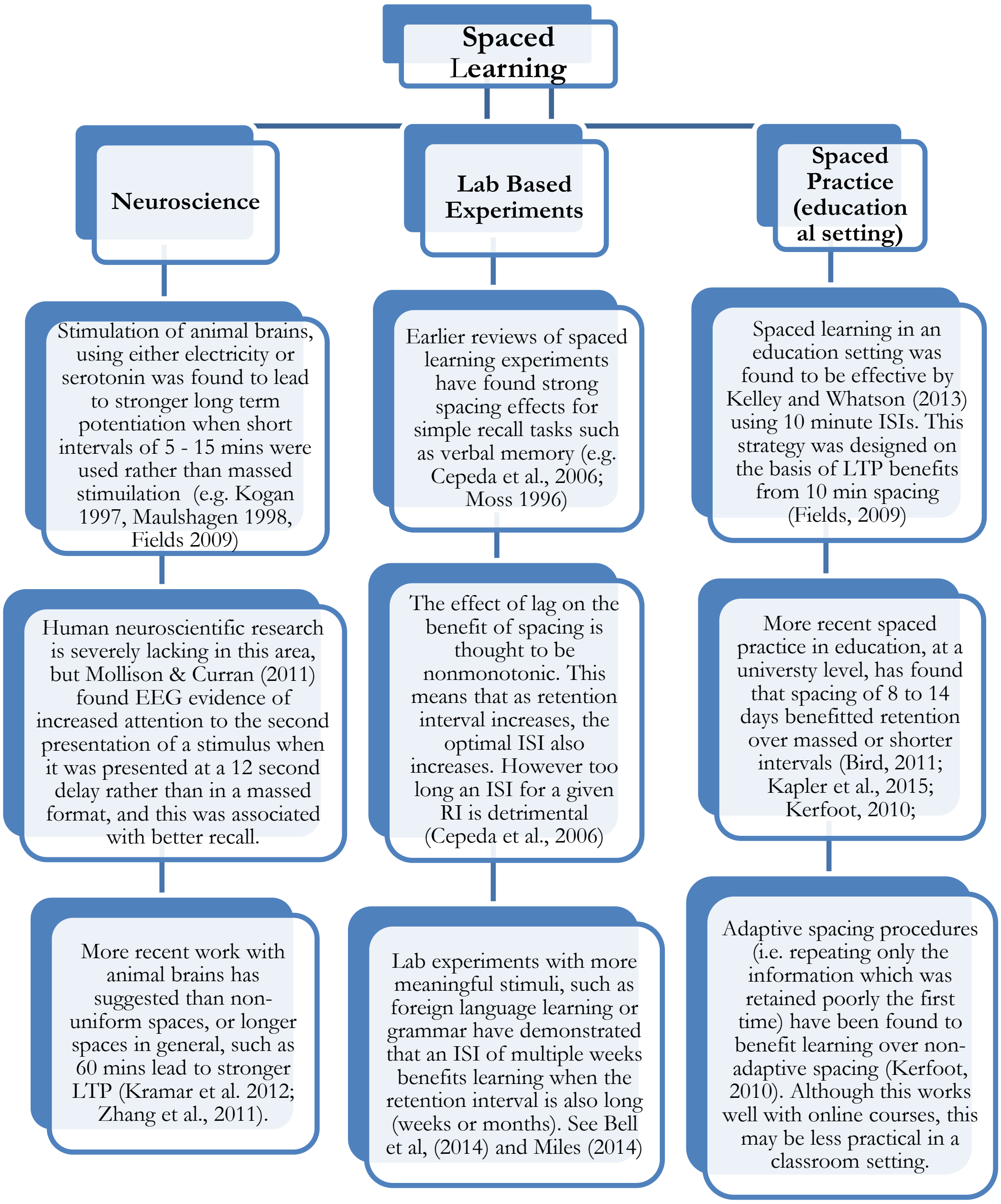
			from sections of rat hippocampus. Studied synaptic facilitation after spaced and massed stimulation (by recording electrical activity and gene presence).	fires itself, is strengthened. Then when it is stimulated again, it produces about 2x as much voltage after 3 bursts with 10 min gaps. This strengthening of synaptic firing is called long term potentiation (LTP). Found that both CREB and a gene called zif268 were present in synapse after spaced stimulation - these are both associated with long term potentiation, and conversion of short to long term memory.
Spaced practice – Online learning course on urology for medical students.	Kerfoot (2010)	Design of online course on complex material (medical school level education)	Randomised controlled trial – Random allocation of blocks of students to either adaptive spaced or nonadaptive spaced education. Adaptive = sent 2 items daily, if correct, then not repeated, if incorrect, repeated 7 days later until all 28 items covered. Nonadaptive = 2 items daily for 40 days, items repeated 10 days later. Retention lag varied due to adaptive period.	Adaptive spaced repetition of items led to students completing 20% fewer items (i.e. more efficient, less unnecessary repetition of known items). Comparable learning with less effort. *This would work well with online course with lots of individual assessment during study – less practical in classroom where lessons need to be tailored to larger numbers.*
Spaced practice – Complex - Native Malay speakers learning English, alongside their degree.	Bird (2011)	Design of grammar and syntax lessons for English as a second language. Administration to university students.	Compared ISIs of 3 and 14 days for retention intervals of 7 and 60 days. Grammar lessons.	No difference in test performance at 7 day RI, between 3 and 14 day ISI. For 60 day RI, only 14 day ISI showed sustained improvement over pre-test. *Paper discusses massed vs distributed, but even the 'massed' was at 3 day intervals. More of a lag effect than spacing effect (which*

				is spaced vs totally non-spaced).
Neuroscience - humans	Mollison & Curran (2011)	EEG research lab.	Comparison of learning of pairs (noun + picture) when presented in spaced (12 second) or massed format. Recorded EEG during task – more similar brain response to spaced items = similar levels of attention leading to better recall	Spaced pairs recalled more accurately than massed. Brain response more similar to second presentation of spaced items than of massed items. This is thought to be indicative of increased attention to P2 when spaced, because less repetition suppression effect.
Neuroscience – animals	Zhang et al. (2011)	Computational modelling of biochemical cascades in animals & Animal research lab.	Used computational modelling to determine optimal ISI for serotonin application to aplysia. Determined that non-uniform applications of serotonin would have optimal effects – ISIs of 10,10,5 then 30 mins. Then tested on aplysia synapses.	Long term synaptic facilitation (i.e. synaptic activity necessary for long term memory, including presence of CREB) was greater for the enhanced non-uniform spacing protocol over standard spacing (5x5 min ISI) or massed (25min constant). *Basically this expands on Maulshagen et al. (1998) by adding the enhanced non-uniform spacing protocol.*
Neuroscience - Humans	Xue et al. (2012)	fMRI lab.	Adults memorised 120 novel faces, half massed, half spaced. Brain activity recorded using fMRI.	If successfully learned, stronger activation in bilateral fusiform gyrus (area associated with face recognition). Massed led to less activity because of repetition suppression (same face in massed suppressed activity). Spacing = less repetition suppression of activity and better recognition.
Neuroscience - Animals	Kramar et al. (2012)	Animal research lab.	Electrical stimulation of rat hippocampal slices at 10, 30 or 60 min intervals.	Twice as much long term potentiation occurred when bursts of electrical stimulation were administered at 60 min intervals than at 10 or 30 min intervals.
Lab based	Jackson,	Basic visuospatial	Younger vs older	Both young and older

experiment with neurobiological component (measurement of amyloid levels in older people).	Maruff and Snyder (2013)	paired learning, plus measurement of amyloid levels in blood.	adults. Each group did both massed and spaced learning of visuospatial information. Spacing intervals were 15 mins. Retention period was 24 hours.	adults showed spacing effect. Subset of older participants showed elevated amyloid levels (a marker of cognitive decline and dementia risk). Showed decrease of spacing effect as amyloid levels increased.
Spaced practice	Kelley and Whatson (2013)	Training of teachers to administer spaced learning techniques in classroom. High stakes test on national curriculum Biology course.	Spaced learning (3 x 20 minutes of stimulation with ISI of 10 mins distraction) vs massed learning (4 months typical teaching).	Spaced learning produced similar long term memory performance to massed learning, despite significantly less teaching time and presence of distractor activities. *Spaced learning is claimed to be much more time effective, i.e. learning per hour of instruction*.
Lab based experiment – complex – learning of new language.	Bell et al. (2014)	Instruction of Swahili-English pairs and subsequent testing.	Participants learned Swahili-English word pairs. Comparison of massed vs spaced (12 hours without sleep) vs spaced (12 hours with sleep) vs spaced 24 hours (with sleep).	All spaced conditions showed spacing effect. 12 hours with sleep and 24 hours with sleep showed enhanced spacing effect over 12 hours no sleep. *It is possible that the longer ISI itself is not solely responsible for enhanced retention. The presence of sleep with a shorter interval may be just as effective.*
Lab based experiment	Huff and Bodner (2014)	Basic lab based learning and memory tasks with undergraduates.	Investigation of the effect of encoding variability (i.e. different tasks when learning lists) on spacing effect.	Related to Raajmakers et al. (2003) Bird et al. 1978 found that encoding variability did not influence spacing effect. This study finds that variable processing did benefit memory, but most strongly when both tasks and type of processing varied across study experiences.
Lab based experiment – complex	Miles (2014)	Teaching of English grammar to South Korean TESOL	Comparison of massed vs spaced (ISIs of 1	Spacing effect found, despite the complex nature of the material

grammar learning.		students.	week then 4 weeks). Retention interval of 5 weeks.	(many lab based experiments are simpler recall tasks).
Spaced practice (undergraduate class)	Kapler, Weston and WIseheart (2015)	Teaching of psychology undergraduates – classroom environment.	Learned material, then reviewed either 1 day or 8 days later. 5 week retention interval.	8 day spaced group scored more highly at the retention test than the 1 day spaced group. *Benefit of longer spacing for longer retention interval and for meaningful/higher level material.*

Figure 3: Summary flow chart of spaced learning research and development in the three areas of neuroscience, experimental psychology and educational practice



Appendix 2: Post-test implementation survey (SMART Spaces Engagement Questionnaire)

Name:______________________________**School:**____________________________________

Class:_______________________________

Please answer the following questionnaire with how much you disagree or agree with the statements based on your experience with Spaced Learning	Strongly Disagree	Disagree	Agree	Strongly Agree
I was enthusiastic to try Spaced Learning	☐	☐	☐	☐
The Spaced learning lessons helped me learn more easily than normal lessons	☐	☐	☐	☐
I think Spaced Learning works well for revision	☐	☐	☐	☐
The spaced Learning lessons ran smoothly	☐	☐	☐	☐
The space activity activities (e.g. juggling) were helpful	☐	☐	☐	☐
Spaced learning would work for all classes, not just revision	☐	☐	☐	☐
I think 3 lessons was enough time for Spaced Learning.	☐	☐	☐	☐
The class as a whole enjoyed Spaced learning	☐	☐	☐	☐
The teacher was confident at delivering Spaced Learning	☐	☐	☐	☐
I felt more motivated to learn during Spaced Learning than in normal classes	☐	☐	☐	☐
I would be happy to try Spaced learning again in the future	☐	☐	☐	☐
I think Spaced Learning would also be useful for other subjects	☐	☐	☐	☐
I found the spaced learning lessons tiring	☐	☐	☐	☐
I found the spaced learning lessons fun	☐	☐	☐	☐
I found the spaced learning lessons helpful for revision	☐	☐	☐	☐
I found Spaced Learning too repetitive	☐	☐	☐	☐
I did not enjoy the space activities (e.g. juggling)	☐	☐	☐	☐

Multiple Choice Section

Please circle one answer for each question

1 How well was the idea behind Spaced learning explained to the class by your teacher?

a. Very well b. Well c. Poorly d. Very poorly

2 Was the activity that you did during the spaces juggling, or some other activity?

a. Juggling b. Some other activity (please specify)____________________________

3 Did the teacher cover all the slides in each lesson or skip over some?

a. Covered all the slides
b. Skipped over some slides in some Spaced Learning lessons
c. Skipped over some slides in every Spaced Learning session

4 Was Spaced learning delivered to your class 3 days in a row?

a. Yes
b. No, there were days in between Spaced Learning lessons
c. No, we received Spaced learning lessons fewer than 3 times

5 Was the time period devoted to juggling, or other activity, the same each day?

a. Yes b. No

6 How were the science subjects divided across the lessons?

a. I had one science subject per day of the project
b. I had all 3 sciences on each day of the project

7 How often did your class take part in the juggling activity?

a. My class juggled multiple times throughout each lesson
b. My class juggled only at the end of the lesson
c. My class performed a different spacing activity throughout each lesson
d. My class performed a different spacing activity only at the end of each lesson

Table 11: Responses to the engagement questionnaire administered to pupils at post-test

	Strongly Disagree	Disagree	Agree	Strongly Agree
	%	%	%	%
I was enthusiastic to try Spaced Learning	5.5%	9.4%	63.3%	21.9%
The Spaced learning lessons helped me learn more easily than normal lessons	6.7%	23.5%	48.6%	21.2%
I think Spaced Learning works well for revision	4.7%	14.4%	51.0%	30.0%
The spaced Learning lessons ran smoothly	5.9%	17.6%	56.5%	20.0%
The space activity activities (e.g. juggling) were helpful	8.2%	23.4%	41.8%	26.6%
Spaced learning would work for all classes, not just revision	12.2%	39.2%	34.1%	14.5%
I think 3 lessons was enough time for Spaced Learning.SL	13.3%	45.3%	33.6%	7.8%
The class as a whole enjoyed Spaced learning	5.6%	20.6%	48.4%	25.4%
The teacher was confident at delivering Spaced Learning	3.1%	5.9%	50.4%	40.6%
I felt more motivated to learn during Spaced Learning than in normal classes	7.0%	25.0%	41.0%	27.0%
I would be happy to try Spaced learning again in the future	7.1%	11.4%	44.9%	36.6%
I think Spaced Learning would also be useful for other subjects	7.9%	18.6%	47.8%	25.7%
I found the spaced learning lessons tiring	18.9%	48.0%	24.4%	8.7%
I found the spaced learning lessons fun	5.1%	12.9%	55.9%	26.2%
I found the spaced learning lessons helpful for revision	5.1%	19.6%	49.8%	25.5%
I found Spaced Learning too repetitive	15.0%	44.9%	29.9%	10.2%
I did not enjoy the space activities (e.g. juggling)	41.6%	36.9%	15.3%	6.3%

Table 12: Responses to the fidelity questionnaire administered to pupils at post-test

Question	Response	%
How well was the idea behind Spaced learning explained to the class by your teacher?	Very Well	41.0%
	Well	50.6%
	Poorly	4.8%
	Very Poorly	3.6%
Was the activity that you did during the spaces juggling, or some other activity?	Juggling	91.5%
	Other	8.5%
Did the teacher cover all the slides in each lesson or skip over some?	Covered all slides	90.2%
	Skipped some slides in some lessons	8.5%
	Skipped some slides in every lesson	1.2%
Was Spaced learning delivered to your class 3 days in a row?	Yes	81.1%
	No, there were days in between lessons	18.5%
	No, we received Spaced learning fewer than 3 times	.4%
Was the time period devoted to juggling, or other activity, the same each day?	Yes	86.4%
	No	13.6%
How were the science subjects divided across the lessons?	One science per day	62.1%
	All three per day	37.9%
How often did your class take part in the juggling activity?	Class juggled multiple times throughout each lesson	68.5%
	Class juggled only at the end of the lesson	27.0%
	Class performed a different spacing activity throughout the lesson	3.3%
	Class performed a different spacing activity only at the end of the lesson	1.2%

Appendix 3: Information Sheet and Memorandum of Understanding for Schools

School Agreement to participate in a feasibility study of the Spaced Learning Project.

School Name and Email --

Name and email of Lead for Project in school---

Aims of the Feasibility Study

The aim of this feasibility study is to develop, and evaluate the materials and resources required for deployment of teaching based on Spaced Learning principles in Y9/10 classes. We hope that the outcomes of this test phase will enable us to develop a method (a programme manual) and materials (for teaching the programme). The results of this feasibility study will contribute towards our understanding of what works in terms of improving retention of science subject knowledge with Y9/10 classes and will be disseminated to schools in England and possibly elsewhere. Ultimately, we hope the results from these test schools will equip us to better understand Spaced Learning and support those wishing to use this teaching and learning style as an additional tool for helping their pupils to retain the subject knowledge required for science attainment at GCSE.

Rationale

The 'Spaced Learning' intervention is a method and materials for teaching that is inspired by models of learning from psychology and neuroscience evidence that indicate spaces between presentation of information improves memory and recall of information. Currently, Spaced Learning takes the form of three bursts of teaching that each cover the same content in different ways, with intervals of distraction activity (e.g. juggling, using clay). The specifics of the model will be explored in feasibility project. For example, there is debate over the optimal length of time between the teaching bursts. However, the intervention will have to be practical for delivery in schools and hence this project will test delivery in real English schools. This project will attempt to explore these issues through a series of research questions.

The research questions

In this study the focus will be on design, feasibility and implementation of a ‘spaced learning’ programme and will therefore test slightly different versions of the programme against each other. There are three main questions regarding the best version of the programme:

1. What version shows best potential efficacy?
2. What version has the best potential for implementation in the classroom?
3. What is the best version for ‘going to scale’ i.e., rolling out across many schools?.

The Methodology

Part 1 – Training

Initially, your teachers will be required to attend meetings and workshops run by the project team to familiarise yourselves with how SL works in practice.

It is envisaged that the project team will work closely with your schools (approx.12-15 schools) in order to ensure that you are given the intervention and are aware of who will participate, what materials you will use, what activities you will try in the spaces, how you will deliver the lesson.

Part 2 – Programme delivery

Schools are asked to attempt to implement different versions of the programme variants in your respective classes (potentially including control groups which will carry on with typical practice.)

Part 3 – Programme data collation for analysis

Your school staff and pupils will be required to provide qualitative and quantitative data for the evaluation team (Queen's University Belfast) to analyse (further details below). Your data will be imperative in adapting the existing materials to further develop the manual for use in future delivery and research of the SLP.

Your participation will help to produce the final output in the form of a 'spaced learning' manual featuring: a literature review; logic model (explicating mechanisms of outcome change); implementation data; training guide; and programme content.

Your participation

1. Your school has been contacted by the project team based on our funder's preferences that schools have significant proportions of pupils eligible for pupil premium and are representative of the population of English state schools in terms of Ofsted ratings, GCSE scores, etc.
2. You also have interest in Spaced Learning principles and a range of familiarity with the concept but have not been implementing spaced learning with those classes participating,

Use of Data

Pupils' test responses and all pupil data will be treated with the strictest confidence. Named data will be matched with the National Pupil Database and shared within the Evaluation team (QUB).

Aggregated and anonymised data will be shared with the Project Delivery team (HTSA) EEF and the UK Data Archive. No individual school or pupil will be identified in any report arising from this initial feasibility study or from the final RCT for the SLP

Responsibilities

The PROJECT TEAM (HTSA) will:

- Organise and deliver training for schools in Spaced learning intervention, making sure teachers understand what is to be done in classes and how SL lessons should be organised
- Provide PowerPoint and other materials to be used
- Be the first point of contact for any questions about the delivery of SL lessons and the use of materials.
- Provide on- going support to schools
- Send out regular updates of the progress of the project in its entirety through a newsletter

The EVALUATION TEAM (QUB) will:

- Randomly allocate schools to variations of the programme
- Collect pre and post GCSE test data from pupils

- Organise and deliver focus groups with pupils and teachers to gather data on the feasibility of implementing a ‘spaced learning’ programme
- Carry out classroom observations of spaced learning lessons
- Be the first point of contact for any questions about the opt-out consent forms, workshops
- Collate and analyse data from these focus groups and lesson observations
- Disseminate research findings

The SCHOOL will

- Ensure the shared understanding and support of all school staff, including the Governing Body, for the project and personnel involved
- Ensure attendance at meetings /workshops organised by the project team
- Consent to random allocation to variants of the programme and commit to the lifetime of the project (September 2015 September 2016) for the successful completion of the initial feasibility study. Remember that some schools or classes could be in the control group and in this case the school will be asked not to deliver spaced learning in any format to these students until after the agreed testing.
- Consent to data collection within the school, in terms of focus groups with pupils and staff (including leadership)
- Consent to data collection within the school, in terms of classroom observations and science measures based on GCSE Science questions.
- Implement interventions as requested based on evaluation team allocation, including if necessary releasing 2-3 staff so they can attend all meetings and workshops
- Agree to changing the way lessons are delivered for the purposes of fulfilling the SL lesson delivery
- Share requested school level data with the Project team
- Inform all parents /carers about the project and collect opt out consent forms
- Be a point of contact for parents /carers seeking more information on the project
- Share requested pupil level data with the project team
- Inform parents /carers about any additional data collection – information provided by Project team

We commit to the Second Feasibility Study (OS) of the Spaced Learning Project

School
Name ____

Headteacher
Name ____

Headteacher Signature ____________________ Date _____________

Names and Emails of Participant staff ____________________________
____

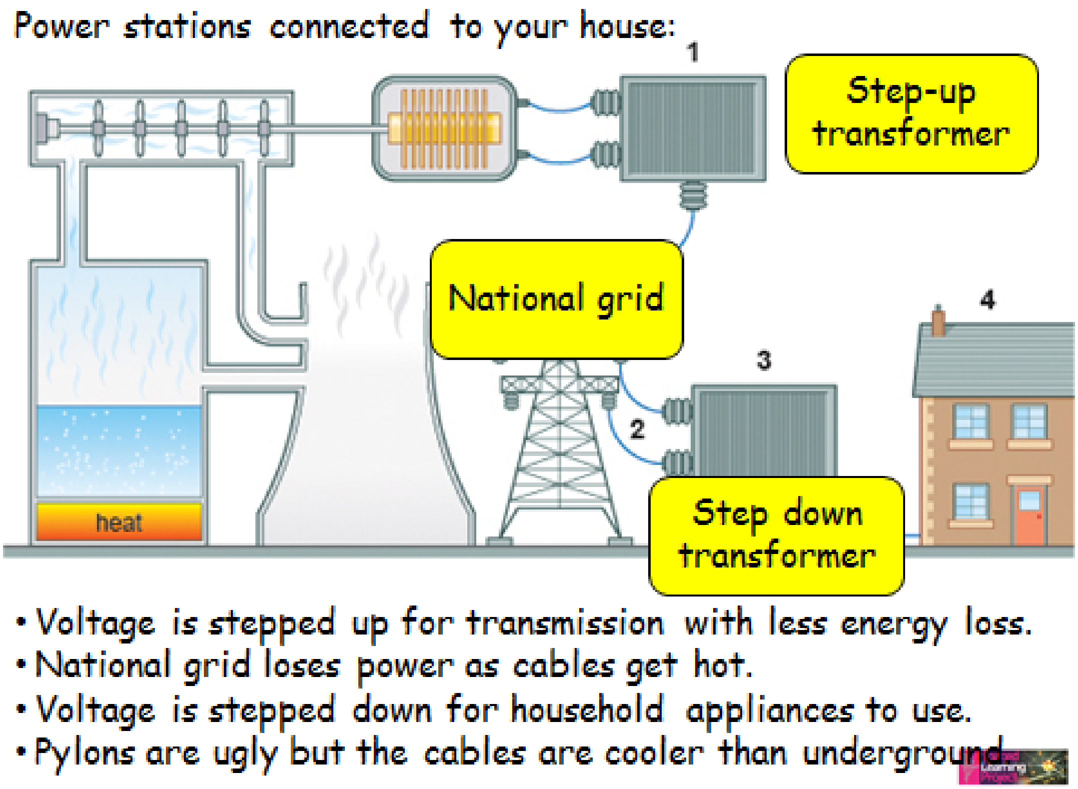
Chair of Governing Body ________________________ Date _________

Thank you for agreeing to take part in this research. ***Please sign both copies, retaining one and returning the second copy to Farhana Zaman at Hallam Teaching School Alliance** Notre Dame **High School, Fulwood Road, Sheffield, S10 3BT***

Appendix 4: Outline of topics, marks and GCSE paper sources for the Science Measure

Subject	Topic	Marks	Source paper	Total marks
Biology Short Answer and Multiple Choice	Keeping Healthy – Balanced diet	1	AQA-BL1FP-Jun14	14
	Metabolic rate	1	AQA-BL1HP-Jun13	
	White blood cells	1	AQA-BL1HP-Jan13	
	Vaccines	2	AQA-BL1HP-Jan13	
	Antibiotics & resistance	2	AQA-BL1HP-Jun14	
	Growing bacteria	1	AQA-BL1FP-Jun13	
	Nervous system – synapses	1	AQA-BL1FP-Jan12	
	Fertility hormones	1	AQA-BL1HP-Jan13	
	Water & salt control	2	AQA-BL1FP-Jan13	
	Plant hormones	1	AQA-BL1FP-Jan12	
	Plant hormones	1	AQA-BL1FP-Jan12	
Chemistry Short Answer and Multiple Choice	Basic atomic model – electron configuration	1	AQA-CHILDHOOd1HP-Jan12	12
	Types of bonding	1	AQchildhoodCH1FP-Jun14	
	Reduction	1	CHILDHOOdA-CH1FP-Jan12	
	Electrolysis	1	AQA-CH1FP-Jun14	
	Phytomining & Bioleaching	1	AQA-CH1FP-Jun14	
	Properties of metals	2	AQA-CH1FP-Jan12	
	Hydrocarbons	1	AQA-CH1FP-Jan12	
	Hydrocarbons	1	AQA-CH1FP-Jan12	
	Cracking alkanes	1	AQA-CH1FP-Jun14	
	Fractional distillation	2	AQA-CH1FP-Jan12	
Physics Short Answer and Multiple Choice	Infrared	1	AQA-PH1FP-Jan13	13
	Change of state	1	AQA-PH1FP-Jan13	
	Convection, conduction, evaporation	2	AQA-PH1FP-Jun13	
	Rate of evaporation	1	AQA-PH1FP-Jan13	
	Energy transfer by design	1	AQA-PH1HP-Jun12	
	U-value & insulation	1	AQA-PH1HP-Jan13	
	Types of energy	3	AQA-PH1FP-Jun14	
	Power stations	2	AQA-PH1HP-Jun12	
	Fossil fuels	1	AQA-PH1FP-Jun13	
Biology Extended Answer	Growing bacteria	6	AQA-BL1FP-Jan13	6
Chemistry Extended Answer	Fractional distillation	6	AQA-CH1FP-Jun13	6
Physics Extended Answer	Va–uums - Insulation	6	AQA-PH1FP-Jun13	6

Appendix 5: Example SMART Spaces slide



Appendix 6: Time lapsed between pre-test and post-test for each school in OS

School	Time lapsed between pre-test and post-test (days)
A	63
B	37
C	65
D	62
E	63
F	63
G	63
H	63
I	49
J	63
K	63
L	65

Appendix 7: Mixed-effects ML regression

Mixed-effects ML regression (IV = pupil engagement, DV = Post-test total score) , Number of pupils n = 223 (from 10 min, 24 hour, 10min/24 hr, and no spaces control groups). Group variable: School Number of schools = 9

Description			
	coef.	**s.e.**	**p=**
Pre-test	**0.85**	**0.04**	**<0.01**
Engagement with SMART SPACES	0.12	0.05	0.01
Constant	**1.65**	**2.20**	**0.45**
-2 Log Likelihood	-671.72		

Mixed-effects ML regression (IV = retention interval, DV = Post-test total score) , Number of pupils n = 313 (all groups). Group variable: School Number of schools = 10

Description			
	coef.	**s.e.**	**p=**
Pre-test	**0.89**	**0.03**	**<0.01**
Retention Interval	-0.02	0.07	0.75
Constant	**5.80**	**1.68**	**<0.01**
-2 Log Likelihood	-945.83		

Appendix 8: Pre, post and gain scores for all variants

		Pre-test		Post-test		Gain
Variant	Score	Mean	SD	Mean	SD	
10 Min	Total score	27.91	9.09	30.82	9.82	2.91
	Short answer	21.25	6.25	23.52	6.59	2.27
	Extended answer	6.66	4.07	7.30	4.41	0.64
24Hour	Total score	25.04	12.81	27.30	14.64	2.26
	Short answer	20.72	9.23	23.19	10.81	2.47
	Extended answer	4.32	4.13	4.12	4.62	-0.20
No Spaces	Total score	21.51	8.54	23.99	8.89	2.48
	Short answer	17.62	6.05	19.35	6.60	1.73
	Extended answer	3.89	3.56	4.63	3.72	0.75
10Min/24Hour	Total score	23.93	10.59	27.55	12.22	3.62
	Short answer	19.93	8.09	22.41	8.69	2.47
	Extended answer	4.00	3.65	5.14	4.30	1.14
No spaces/materials	Total score	25.28	7.16	26.82	9.80	1.54
	Short answer	19.06	5.41	20.32	6.94	1.26
	Extended answer	6.23	2.88	6.50	3.90	0.27
No spaces/materials excluding school 'L'	Total score	23.62	6.51	24.07	7.52	0.45
	Short answer	17.64	4.81	18.34	5.38	0.7
	Extended answer	5.98	2.80	5.73	3.38	-0.25

This document is available for download at **www.educationendowmentfoundation.org.uk**

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk