Retrieval practice in primary science lessons

Craig Westby and Carl Badger

Old Hill Primary School and Nursery

November 2018

About IEE Innovation Evaluation Grants

The first four IEE Innovation Evaluation Grants were awarded in February 2017. Funded by the Institute for Effective Education (IEE), these grants supported pilot evaluations of innovations of teaching and learning approaches based on the Research Schools Network's goal of improving the attainment of pupils by increasing the use of evidence-based practices.

Since then a further 26 projects have been successful in their application for an IEE Innovation Evaluation Grant, bringing the total number to 30. The applications we received included a wide range of interesting, school-led innovations – from after-school film clubs to improve the creative writing of Year 5 pupils, to the use of audio feedback with Year 12 pupils – and we were really impressed with the thought that applicants had put into how these innovations could be evaluated.

The evaluations are small-scale, and test the kinds of innovations that schools are interested in. This is very much a "bottom-up" exercise, allowing schools to get some indicative evidence behind real-world initiatives. Many evaluations are now coming to an end, and we are starting to publish reports on the findings. It is important remember that these are small-scale projects, often carried out in one school, so it is not possible to generalise their findings. In fact, the main benefit of the Innovation Evaluation projects may be in the process, rather than the findings.

Contents

Executive summary

Key Stage 2 science sampling tests reveal a significant decline in attainment since national tests were abolished in 2009/10, with only 23% achieving an estimated expected standard in the 2016 samples compared to 88% achieving a level 4 in 2009 (*Key Stage 2 science sampling 2016 Methodology note and outcomes July 2017, Standards & Testing Agency*). The gap between pupil premium and non-pupil premium pupils is also significant., we wanted to test the effects of retrieval practice in primary science lessons, in the form of multiple-choice quizzing (MCQ), to examine how well pupils retain the content they have been taught.

Intervention pupils were assigned to answer multiple-choice questions using the quizzing app, Socrative, at the beginning of science lessons. Each 10-minute quiz tested pupils' knowledge and application of information learnt from the previous lesson(s). Control groups were asked to reread materials from previous lessons for 10 minutes. Socrative quiz questions, test questions and reading materials were aligned to the National Curriculum objectives for each unit of work studied.

Summary of the evaluation

For the evaluation, 188 pupils were selected from Year 2–5 classes in two urban schools: Old Hill Primary and St James' CofE Primary. Both schools are based in Sandwell and are comparable, with half the pupils eligible for pupil premium. Classes were randomly assigned to either intervention or control group. Pupils undertook a unit pre-test for baseline percentage measures and two further delayed tests once the unit of work was completed: after one week and 12 weeks respectively. Additionally, pupils and teachers completed a simple survey before and after the innovation, to measure attitudinal responses to science lessons, subject knowledge and tests.

Summary of findings

Intervention pupils attained an overall effect size of +0.26 for the one-week delayed test. More interesting, was the 12-week delayed intervention effect size of +0.58. The results suggest that pupils who practise multiple choice quizzing not only perform better in a one-week delayed test than their peers, but perform considerably better after 12 weeks, thus demonstrating a greater retention of learning when retrieval practice is employed in lessons. Results for pupil premium pupils are particularly of interest, revealing no loss of learning in Years 2 and 5 and only a 5% percentage loss in Year 3.

Introduction

The problem

In 2010, the end-of-Key-Stage-2 science SATS were abolished, although sampling tests in Year 6 continue. The table below indicates that the national standard in science is low; the gap between children eligible for free school meals (FSM) and non-FSM children is significant and results are declining.

TABLE 1: ESTIMATED PERCENTAGE OF PUPILS ACHIEVING THE EXPECTED STANDARD BASED ON KEY STAGE 2 SCIENCE SAMPLING ASSESSMENTS IN 2014 AND 2016

	95% confidence interval in 2014	Estimated percentage achieving expected standard in 2014	Estimated percentage achieving expected standard in 2016	95% confidence interval in 2016
All pupils	28.06	26.82–29.31	22.77	21.61–23.92
Boys	27.86	26.03–29.68	22.73	21.02–24.45
Girls	28.28	26.26–30.31	22.80	20.80–24.80
FSM	13.14	10.75–15.53	9.00	6.80–11.19
Non-FSM	31.35	29.93–32.76	25.28	23.99–26.58
EAL	23.17	20.25–26.09	17.41	14.64–20.17
Non-EAL	29.06	27.63–30.49	23.77	22.49–25.04

(Key Stage 2 science sampling 2016 Methodology note and outcomes July 2017, Standards & Testing Agency)

On a local scale, we have evidence from end-of-year science assessments in Years 3–6 that mirror these national figures. One reason for this might be that pupils are not being tested enough throughout the year and teachers have not provided pupils with opportunities for retrieval practice. Consequently, pupils are beginning secondary school with a lack of scientific knowledge of the primary curriculum because they have forgotten the content and therefore need to play ‘catch up’ to stay in touch with the already demanding curriculum. We hypothesised that pupils who have regular opportunities for retrieval practice will better retain information in the long-term, enter secondary school with more embedded knowledge, and will therefore be more successful.

Review of existing research

There is strong evidence that demonstrates testing is a powerful technique to enhance learning: the act of retrieving information from memory promotes the ability to recall material again in the future (Carpenter & DeLosh, 2005;Carrier & Pashler, 1992). Evidence from a number of studies reveals that retrieval practice in authentic classroom settings improves long-term learning (McDaniel et al., 2011). The study, *The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist* (Agarwal et al., 2012) showed that 6th–8th graders (Years 7–9) who undertook multiple-choice quizzes performed better in exams. Further evidence shows that practising retrieval improves learning compared to rereading information (Roediger and Karpicke, 2006). Research has also shown that the difficulty of initial retrieval is correlated with later retention (Karpicke & Roediger 2007; Benjamin, Bjork, & Schwartz, 1998), along with direct evidence that delaying an initial retrieval attempt enhances performance on a later end-of-unit test (Jacoby, 1978; Whitten & Bjork, 1977).

Description of the innovation

Pupils in the intervention group were trained in using the Socrative app to answer multiple choice questions for 10 minutes at the beginning of each science lesson to test their retrieval skills. Control groups experienced the same content but re-read materials, textbooks or notes pages, in place of the multiple choice quiz. Questions were aligned to the National Curriculum objectives for each unit of work studied and pupils were tested on their retention at two intervals after the science unit had been completed.

Research question

What impact will retrieval practice, in the form of multiple-choice quizzing (6 x 10-minute sessions), during the teaching of one science unit have on performance in delayed science tests for pupils in Years 2–5?

Within this question we also examined:

1. The difference between the one-week delayed test and the 12-week delayed test percentage scores for both intervention and control groups.
2. If retrieval practice could potentially help to close the gap between pupil premium and non-pupil premium children.
3. Pupil attitudes towards testing.

Method

Sample

Two urban West Midlands schools took part in the evaluation, both based in Sandwell, a district with a high level of deprivation. The lead school, Old Hill Primary, is slightly smaller-than-average with a large majority of pupils being White British. The proportion of pupils (49%) who are disadvantaged and receive support from the pupil premium is well above average. St James CofE Primary is a larger-than-average-sized primary school, where most pupils are of White British heritage. The proportion of pupils eligible for the pupil premium (50%) is similar to Old Hill Primary, again much higher than the national average. Both schools are below national figures for SEND, including those with a statement.

All pupils were selected to participate in the study except those who were performing at well below national expectations for science or who were unable to access the tests, at the discretion of the teacher. Pupils who missed more than two teaching sessions were not included in results. A total of 188 pupils took part in the innovation, of which 103 were non-pupil premium and 85 pupil premium.

Allocation to groups

Groups were allocated via a random coin toss and were as follows:

TABLE 2

Year groups	Old Hill Primary	St James' CofE Primary
Y2	Intervention	Control
Y3	Intervention	Control
Y4	Control	Intervention
Y5	Intervention	Control

Details of the groups are shown below.

TABLE 3

	Year 2 control	Year 2 intervention
sample size (N)	**23**	**21**
non-pupil premium	17	14
pupil premium	6	7

	Year 3 control	Year 3 intervention
sample size (N)	**23**	**26**
non-pupil premium	17	11
pupil premium	6	15
	Year 4 control	**Year 4 intervention**
sample size (N)	**22**	**24**
non-pupil premium	9	12
pupil premium	13	12
	Year 5 control	**Year 5 intervention**
sample size (N)	**24**	**25**
non-pupil premium	10	13
pupil premium	14	12

Description of the innovation

A baseline test was completed by pupils before studying the science unit of work as follows:

- Year 2: Living things and their habitats
- Year 3: Rocks
- Year 4: States of matter
- Year 5: Properties and changes of materials

Pupils in the four randomly-assigned intervention groups were trained in using the Socrative app to answer multiple-choice questions, written by the class teacher, for 10 minutes at the beginning of each science lesson. It was decided that at week three of the trial (session two), the pupils would only be given five minutes as opposed to 10 minutes due to lack of content taught at that stage. After each answer, the Socrative app would reveal the correct answer with an explanation if appropriate. Pupils continued to repeat the quiz during the allotted time. No scoring system was assigned to the task. Teachers were permitted to read questions on the screen but not allowed to ask retrieval-based questions based on previously taught content eg, “Do you remember last week when we looked at…”.

Four control groups were allocated a re-reading task, which consisted of notes from previous lesson(s) or re-reading textbook materials, for 10 minutes at the beginning of the lesson. Teachers were permitted to answer questions but not allowed to ask retrieval-based questions based on previously taught content.

Teachers in both groups were given unit overview guidance relating to which national curriculum objectives to cover along with suggested aspects to focus on. Teachers in both groups were then free to teach content as usual but were requested not to ask any retrieval questions in relation to content from previous lessons. Over the course of the unit, as the

content increased, so did the materials – Socrative quiz questions for intervention or notes/text for control groups. For example, at week three the pupils would have only had a small amount of content to quiz or re-read whereas by week nine they would have the opportunity to quiz or re-read almost a whole unit of work (see appendix 1, intervention timeline).

Training, ongoing support and educational resources

Training consisted of one staff meeting delivered to both schools setting out expectations and timescales of the innovation, prior to commencement. The innovation lead met separately with teachers of both the intervention and control groups to explain process measures and clarify any points teachers were unclear of. The lead also took this opportunity to train staff who were new to the Socrative app as well as showing example question styles that could be used. Science textbooks were purchased for St James' CofE Primary to match those used in Old Hill Primary to ensure quality and consistency across the schools.

Teachers were provided with:

- an overview of their science unit with suggested timings
- a class register to record pupil attendance in the unit
- an observation schedule: 'Do & Don't' checklist
- week-by-week guidance of the process
- pupil questionnaires
- teacher questionnaires.

Monitoring visits to the intervention and control groups were carried out to ensure consistency of delivery during the retrieval practice and re-reading stage of the lesson (the first five- to 10-minutes of the science lesson).

Outcome measures

A total of three tests were administered to pupils during the innovation: a baseline test seven days before teaching commenced (Test A); a one-week delayed test post unit completion (Test B); and a 12-week delayed test post unit completion (Test A). After examining many commercially produced tests for science, we deemed them not satisfactory to measure the outcomes of the innovation due to: poor wording; a lack of unit coverage; or not enough application/knowledge questions. Therefore, tests were produced in line with National Curriculum objectives and expectations, similar to past Standards and Testing Agency-produced end of Key Stage tests. Teachers administered the tests to their pupils under strict test conditions.

To avoid bias, pupils were assigned a randomly-generated number to ensure both school and pupil anonymity. Test scripts were marked by staff not involved in the innovation using a predetermined mark scheme to avoid ambiguity.

Process evaluation

The lead researcher and deputy headteacher from St James' CofE Primary monitored both schools (together and separately) on two occasions to ensure consistency for both intervention and control groups. The following observation criteria was used:

TABLE 4

Intervention		Control	
Do	Don't	Do	Don't
Insist on silence in the classroom. Support the reading of words. The teacher must minimise and close the teacher feedback screen. Where at all possible pupils should work independently After the 10-minute Socrative quiz, new learning should commence. Science must only be taught in the 90-minute lessons.	Pupils must not discuss questions or answers with each other. Ask retrieval-based questions to pupils from previous lessons (Can you remember from last week, what the function of the roots are?). Remove children from lessons (if any pupil needs to leave the lesson, record on your science register).	Insist on silence in the classroom Support the reading of words. Allow pupils to re-read notes/materials that already have been studied. Where at all possible pupils should work independently. After 10 minutes re-reading, new learning should commence. Science must only be taught in the 90-minute lessons.	Pupils must not discuss questions or answers with each other. Ask retrieval-based questions to pupils from previous lessons (Can you remember from last week, what the function of the roots are?). Remove children from lessons (if any pupil needs to leave the lesson, record on your science register).

Teachers naturally like to ask their pupils questions, and teachers from both intervention and control groups spoke of how difficult it was to teach without inadvertently asking retrieval questions. They found this happened within science lessons and then also when making links to the science unit when teaching other subjects. Although conditions in science lessons were monitored, it would be impossible to eradicate all learning that may have affected retention in the science units.

Pupils and teachers completed a Likert-type survey before the pre-test and repeated it at week 20, before the 12-week delayed test, to explore changes in anxiety and attitude towards knowledge and testing. The surveys for pupils included rating scales for pupil self-perception and reflection about testing, as well as anxiety levels. The pupil survey can be seen in appendix 2. For teachers, the surveys consisted of rating scales relating to teaching methods, workload, pupil retention and pupil motivation. The teacher survey can be seen in appendix 3.

Data analysis

Outcome data

A percentage score was calculated for each pupils' performance in the pre-test, one-week delayed test and 12-week delayed test. These were used to calculate the mean percentage

score on each test for the intervention group and for the control group in each year. Mean percentage points in each of the three tests was also calculated for pupil premium and non-pupil premium pupils in the intervention and control group in each year.

Effect sizes were calculated for each year group by comparing the progress of the intervention and control group between:

1. The pre-test and one-week delayed test.
2. The pre-test and 12-week delayed test.

An overall effect size was calculated by comparing the progress of all pupils in the intervention group and all pupils in the control group between:

1. The pre-test and one-week delayed test.
2. The pre-test and 12-week delayed test.

Process evaluation data

Pupil responses to survey questions were scored as follows: three points for a smiley face, two points for a neutral face, one point for a sad face. Median responses to each question were calculated for the intervention and control group in each year.

Teacher responses to survey questions were scored as follows: five points for strongly agree, four points for agree, three points for no opinion, two points for disagree, one point for strongly disagree. Median responses to each question were calculated intervention group and control group teachers across year groups.

Cost analysis

The overall production costs of the tests, from formulation, to printing, were larger than was originally expected. It was incredibly time-consuming to produce quality questions over four different units of work. Originally, we had expected to purchase published materials, but upon examination, they were not fit for the purpose of this innovation. Actual costs of running multiple-choice online quizzes can be free. Currently, Socrative and Kahoot – both online platforms to present MCQ – are free, depending on the plan, mainly consisting of how many classes a teacher wants to work with. The use of summative tests was integral to the innovation to help measure impact, however to reduce costs, schools may well decide not to use summative tests within their setting as they are an unnecessary commodity in implementing this intervention.

The largest expenses incurred related to the evaluation costs inherent in a research project of this kind: monitoring conditions, staff training, data analysis and report writing. With visits to two sites, staff training was not as efficient as it could have been. However, to ensure staff were clear on how the project would affect their usual teaching as well as training them in conditions, staff were met one-to-one. If the innovation was to expand further, procedures would need to be arranged to enable all staff to be trained simultaneously.

TABLE 4: TOTAL COST OF PROJECT

Budget item	Amount
Production of test A and B for x4 units of work	£2,200
Socrative Pro	£600 (£300 per school)
Staff training (including project process evaluations and conditions)	£3,000 (over both schools)
Release time for monitoring conditions	£2,000 (over both schools)
Data analysis and report writing	£2,200
Total expenditure	**£10,000**

TABLE 5: COST OF DELIVERING THE INTERVENTION

Budget Item	Amount
Socrative Pro	£300
Staff training for Socrative	£800
Total expenditure	**£1,100**

For a one form-entry primary of approximately 210 pupils, a cost of £5.23 per pupil, presuming schools have current ICT equipment that can present Socrative quizzes online, has been estimated for delivering the intervention. However, schools may decide to use the free Socrative plan to begin with, which would lower the cost to £3.80 per pupil.

Results

Outcome measures

All year groups

Intervention pupils attained an overall effect size of +0.26 for the one-week delayed test, and an effect size of +0.58 at 12 weeks. These effect sizes were calculated for the whole sample by comparing progress of all participating pupils between the pre-test and the one-week delayed test, and between the pre-test and 12-week delayed test.

Year 2

TABLE 6: YEAR 2: **EFFECT SIZE**

	Mean pre-test percentage	Mean one-week delayed test percentage	Mean 12-week delayed test percentage	Effect size one-week delayed test	Effect size 12-weeks delayed test
Control (N=23)	19	49	35	0.885	1.393
Intervention (N=21)	30	75	76		

TABLE 7: YEAR 2: PERCENTAGE POINT INCREASE/DECREASE

	Pre-test % to one-week delayed test	One-week delayed % to 12-week delayed test
Year 2 control	+30	-14
Year 2 intervention	+45	+1

CHART 1: YEAR 2: OVERALL

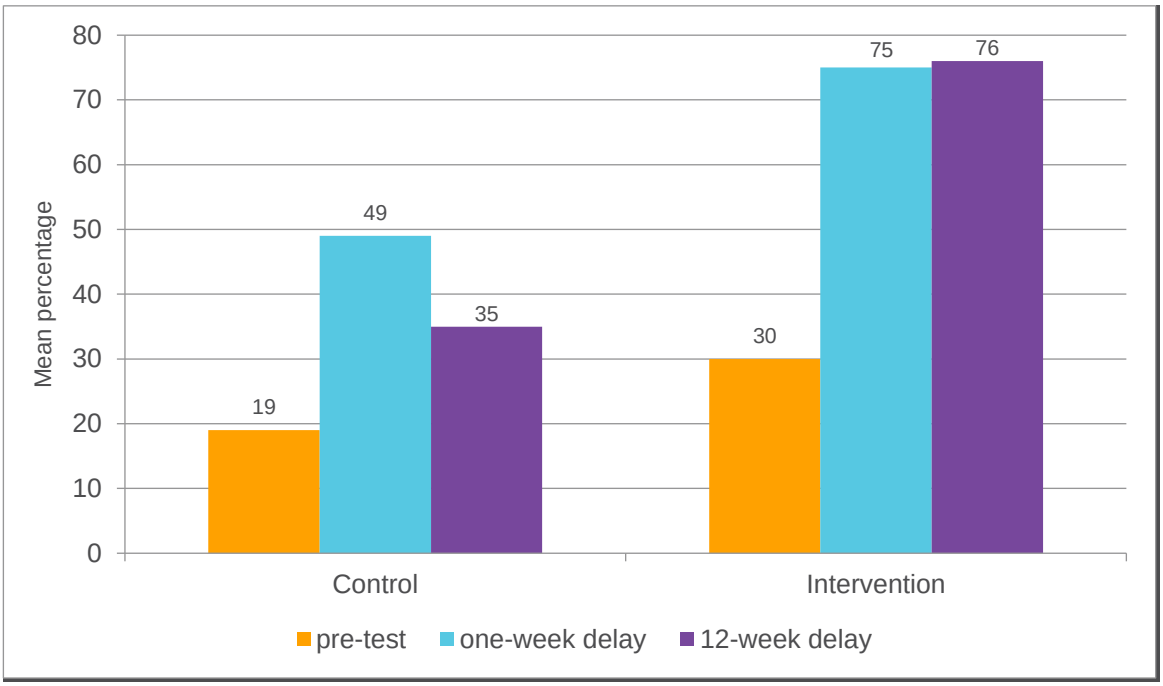
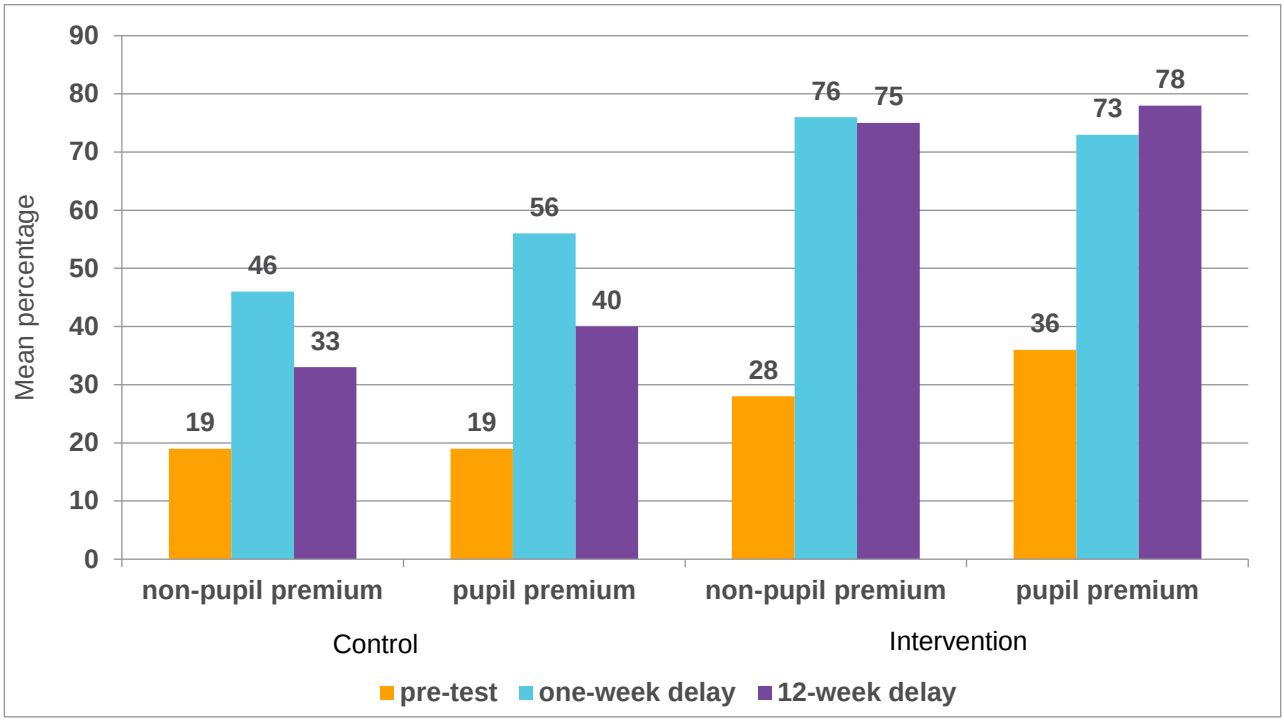


TABLE 8: YEAR 2: PUPIL PREMIUM

	Year 2 control	Year 2 intervention
sample size (N)	**23**	**21**
Non-pupil premium	17	14
pupil premium	6	7

CHART 2: NON-PUPIL PREMIUM VS PUPIL PREMIUM



Results from Year 2 reveal a decisive outcome for the intervention group in not only the one-week delayed test but, more importantly for this study, the 12-week delayed test. The intervention group were able to retain all of what was learnt after the science unit whereas the control group, although making a mean percentage increase of 30 from the pre-test to one-week delayed test, 'lost' 14 percentage points of their learning after 12 weeks: falling from mean = 49% to mean = 35%.

Pupil premium data for Year 2 also shows positive results for the intervention group with pupil premium pupils out-performing all sub-groups.

Year 3

TABLE 9: YEAR 3: EFFECT SIZE

	Mean pre-test percentage	Mean one-week delayed test percentage	Mean 12-week delayed test percentage	Effect size one-week delayed test	Effect size 12-week delayed test
Control (N=23)	19	49	35	0.214	0.807
Intervention (N=26)	29	62	57		

TABLE 10: YEAR 3: PERCENTAGE POINT INCREASE/DECREASE

	Pre-test % to one-week delayed test	One-week delayed % to 12-week delayed test
Year 3 control	+30	-14
Year 3 intervention	+33	-5

CHART 3: YEAR 3: OVERALL

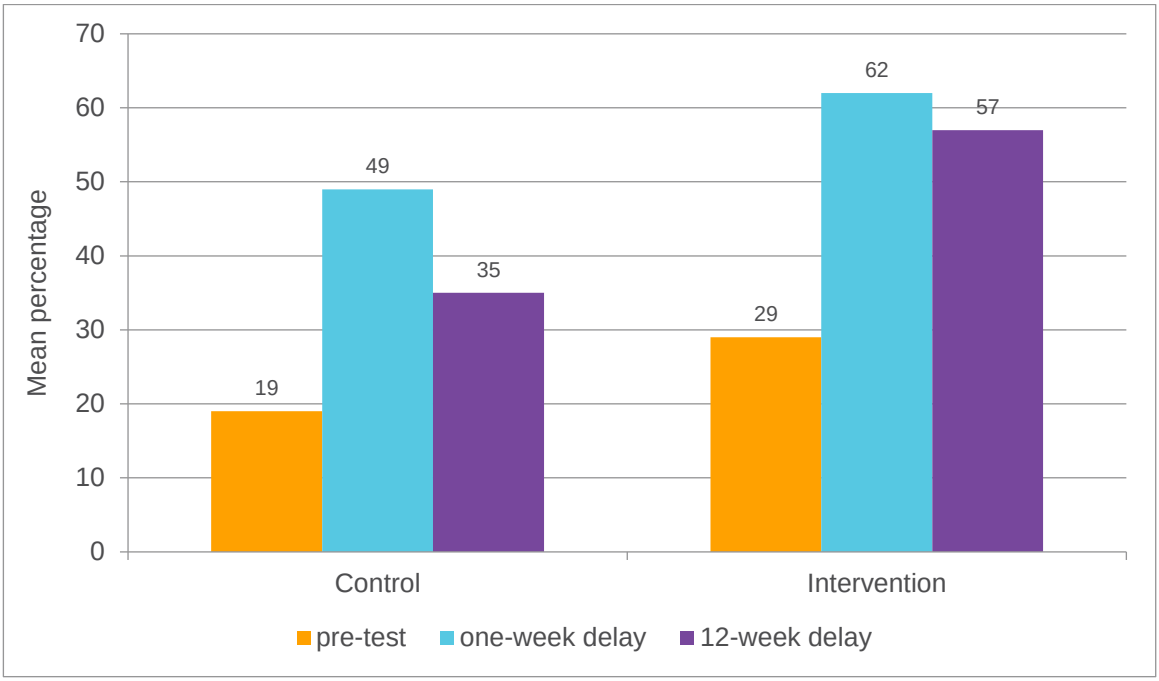
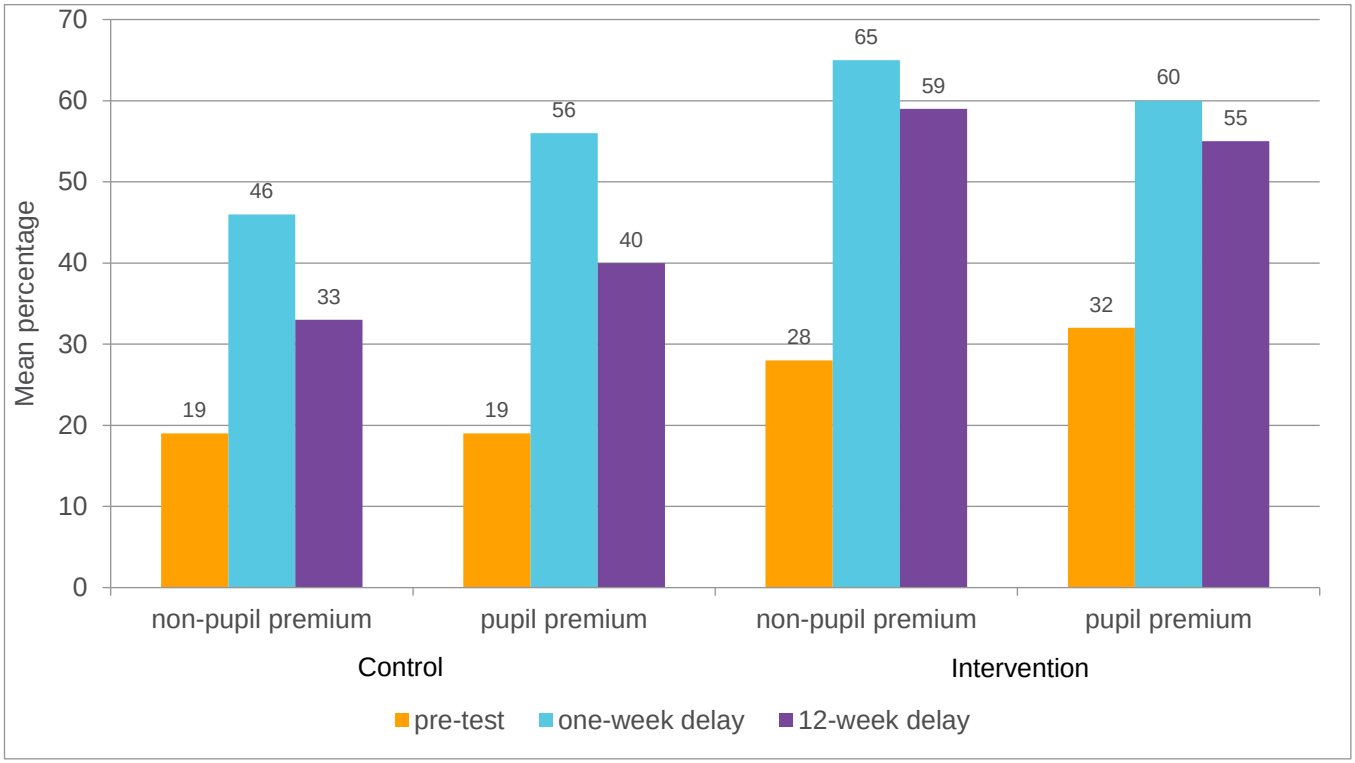


TABLE 11: YEAR 3: PUPIL PREMIUM

	Year 3 control	Year 3 intervention
sample size (N)	**23**	**26**
non-pupil premium	17	11
pupil premium	6	15

CHART 4: NON-PUPIL PREMIUM VS PUPIL PREMIUM



Year 3 results paint a similar picture with the intervention group showing only a five percentage point mean decrease from the one-week delayed test to the 12-week delayed test. In comparison, outcomes for the control group were lower on each test, significantly a 14 percentage point mean decrease from the one-week delayed test to the 12-week delayed test. Pupil premium pupils performed in-line with their peers.

Year 4

TABLE 12: YEAR 4: EFFECT SIZE

	Mean pre-test percentage	Mean one-week delayed test percentage	Mean 12-week delayed test percentage	Effect size one-week delayed test	Effect size 12-week delayed test
Control (N=22)	36	65	43	-0.600	0.002
Intervention (N=24)	39	55	46		

TABLE 13: YEAR 4: PERCENTAGE POINT INCREASE/DECREASE

	Pre-test % to one-week delayed test	One-week delayed % to 12-week delayed test
Year 4 control	+29	-22
Year 4 intervention	+16	-9

CHART 5: YEAR 4: OVERALL

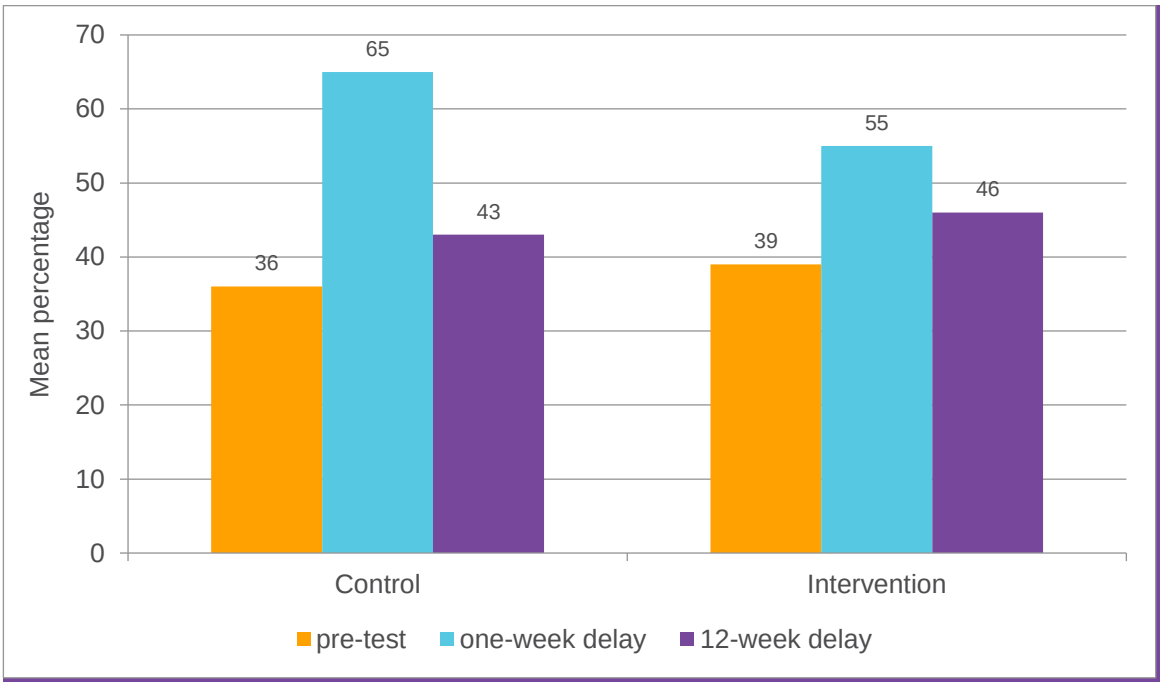
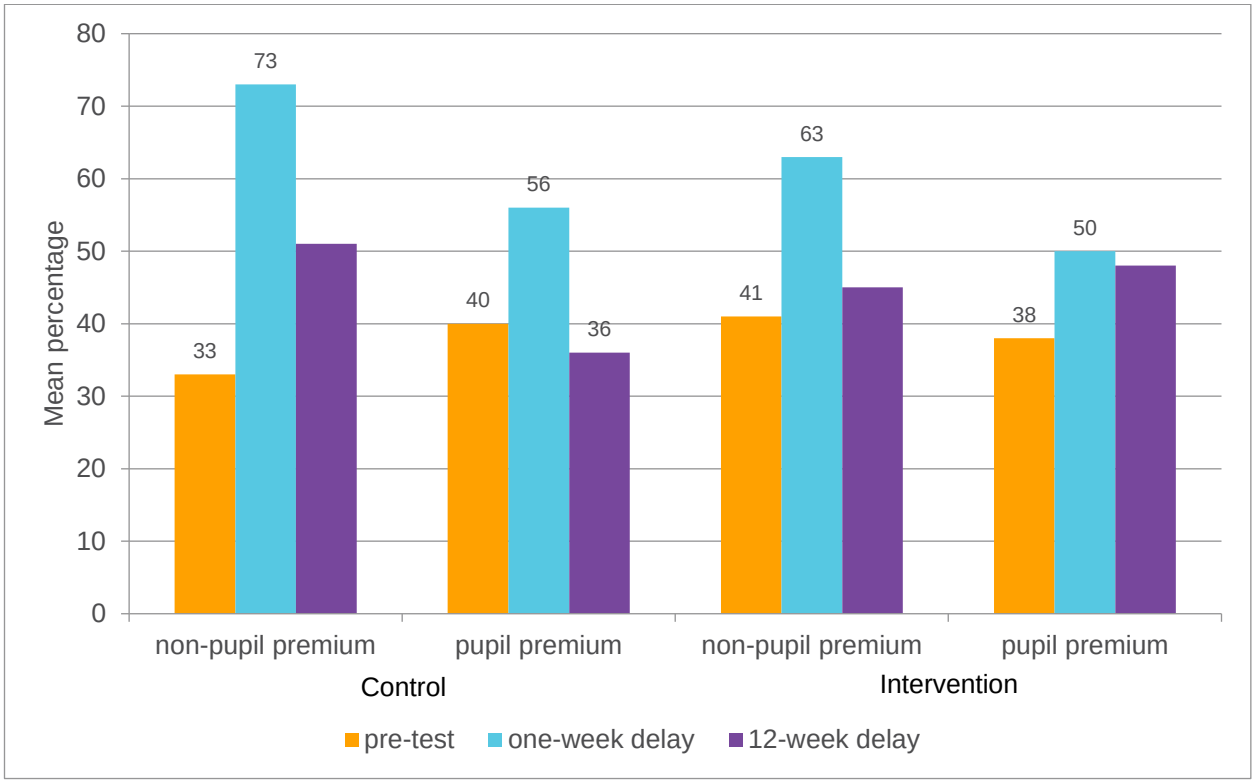


TABLE 14: YEAR 4: PUPIL PREMIUM

	Year 4 control	Year 4 intervention
sample size (N)	**22**	**24**
non-pupil premium	9	12
pupil premium	13	12

CHART 6: NON-PUPIL PREMIUM VS PUPIL PREMIUM



Results for Year 4 demonstrate that the control group performed 10% better in the one-week delayed test. Both groups' baseline score revealed similar starting points, yet interestingly, the control group 'lost' a significant 22 of their mean percentage points from the one-week delayed test to the 12-week delayed test, compared with a nine percentage point mean decrease for the intervention. Consequently, although the control group out-performed the intervention in the one-week delayed test, the intervention group performed slightly better after 12 weeks, thus indicating the quiz aided learning retention. Pupil premium pupils in the intervention group also scored 12% higher after the 12-week delayed test compared with pupil premium pupils in the control group.

Unfortunately, the intervention group at St James' CofE Primary were unable to access the multiple-choice quiz in its intended form (either on iPads or on desktop computers) for four sessions out of six, due to server technical problems. Consequently, a paper version was used with the same questions. However, the class teacher reported that the pupils were not as focused, could not repeat the questions and received no instant feedback, in this unintended format.

Year 5

TABLE 15: YEAR 5: **EFFECT SIZE**

	Mean pre-test percentage	Mean one-week delayed test percentage	Mean 12-week delayed test percentage	Effect size one-week delayed test	Effect size 12-week delayed test
Control (N=24)	39	40	57	0.847	0.133
Intervention (N=25)	49	66	68		

TABLE 16: YEAR 5: PERCENTAGE INCREASE/DECREASE

	Pre-test % to one-week delayed test	**one-week delayed % to 12- week delayed test**
Year 5 control	+1	+17
Year 5 intervention	+17	+2

CHART 7: YEAR 5: OVERALL

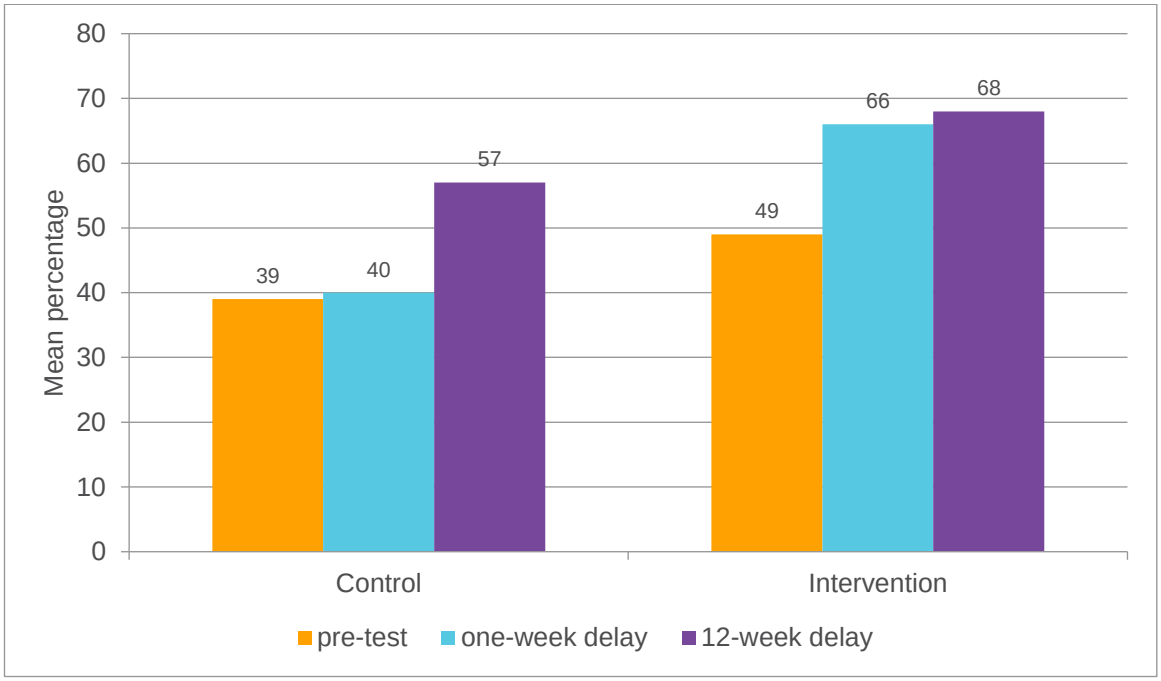
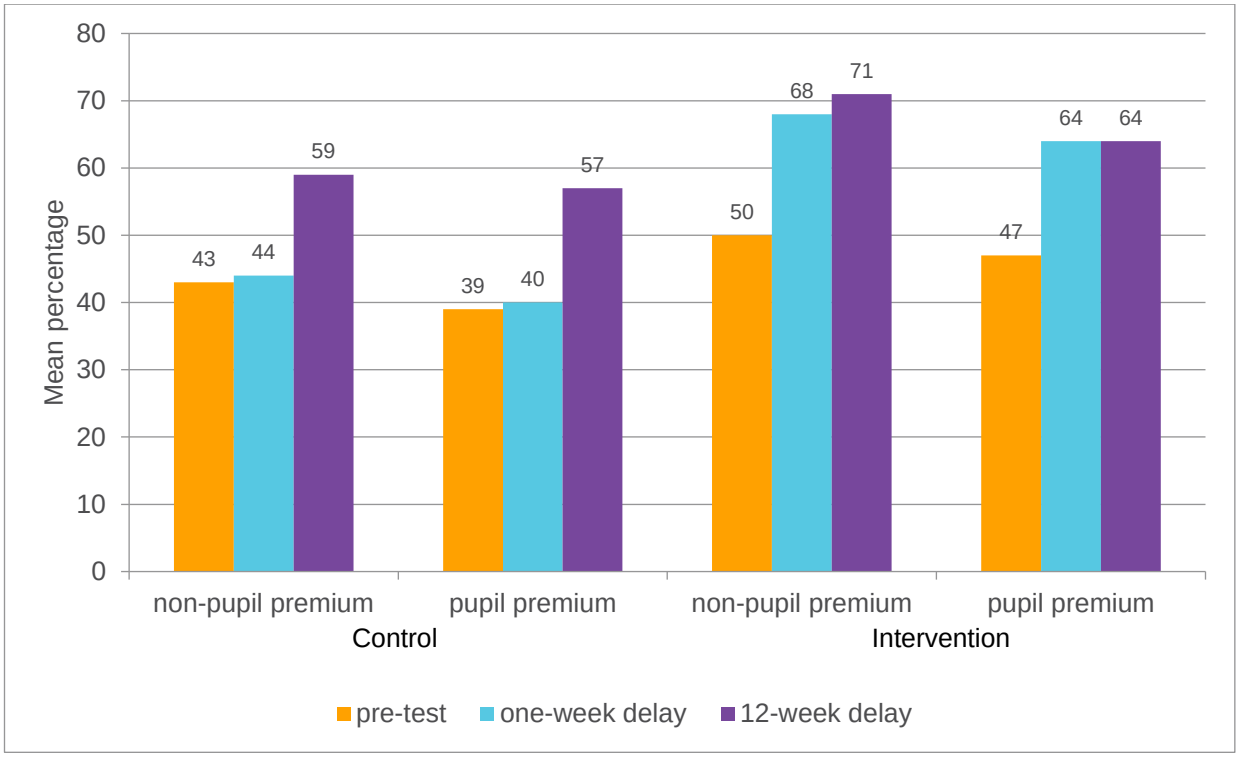


TABLE 17: YEAR 5: PUPIL PREMIUM

	Year 5 control	Year 5 Intervention
sample size (N)	**24**	**25**
non-pupil premium	10	13
pupil premium	14	12

CHART 8: NON-PUPIL PREMIUM VS PUPIL PREMIUM



Year 5 data reveals unusual findings for the control group. Whereas a forgetting curve, such as Ebbinghaus (1885), would display an increase in memory and outcomes during and immediately after studying a unit before declining, the control showed only a one percentage point increase from baseline to the one-week delayed test but then a 17 percentage point increase from the one-week delayed to the 12-week delayed test. Although the control group teacher assured us that they followed the innovation protocol, data would suggest that the Year 5 control class received teaching in this science unit at some point between the one-week and the 12-week delayed post-tests. Evidently, this affects the overall Year 5 effect size of the retrieval intervention. Yet, despite this, it is apparent that retrieval practice has enabled pupils to remember what they have learnt after 12 weeks, for both non-pupil premium and pupil premium pupils in the intervention group.

Process evaluation

Pupil survey

Median scores for intervention and control group pupil responses to the survey in each year are presented below.

3 =  2=  1 = 

TABLE 18: YEAR 2

		Do you enjoy science?	Are you good at science?	Do you know lots of stuff about science?	Can you remember lots of science stuff from last year?	Are you good at answering science questions?	How do tests make you feel?
Intervention	Pre-test	3	2	2	2	3	3
	12-week post unit	3	2	3	1	2.5	3
Control	Pre-test	3	2	3	3	2	3
	12-week post unit	3	2	2	2	3	3

TABLE 19: YEAR 3

		Do you enjoy science?	Are you good at science?	Do you know lots of stuff about science?	Can you remember lots of science stuff from last year?	Are you good at answering science questions?	How do tests make you feel?
Intervention	Pre-test	3	2	2	1	2	2
	12-week post unit	3	2	2	1	2	2
Control	Pre-test	3	2	2	2	2	3
	12-week post unit	3	2	2	2	2	2

TABLE 20: YEAR 4

		Do you enjoy science?	Are you good at science?	Do you know lots of stuff about science?	Can you remember lots of science stuff from last year?	Are you good at answering science questions?	How do tests make you feel?
Intervention	Pre-test	3	2	2	2	2	2
	12-week post unit	3	2	2	2	2	2
Control	Pre-test	3	2	2	2	2	2
	12-week post unit	3	2	2	2	2	2

TABLE 21: YEAR 5

		Do you enjoy science?	Are you good at science?	Do you know lots of stuff about science?	Can you remember lots of science stuff from last year?	Are you good at answering science questions?	How do tests make you feel?
Intervention	Pre-test	3	2.5	2	1.5	2	3
	12-week post unit	2.5	2.5	2	2	2	3
Control	Pre-test	3	2	2	2	2	2
	12-week post unit	3	2	2	2	2	2

On the whole, pupils were largely positive about science in both intervention and control groups, and at both time points. Intervention and control groups in the same year group tended to respond alike to the questions. Similarly, there was little variation in the median response over time. Furthermore, pupils in intervention groups did not feel more negatively about tests than control group peers.

Teacher survey

TABLE 22: CONTROL PRE-TEST

	I enjoy teaching science.	Science should be taught through a themed topic alongside other subjects.	Pupils should be taught scientific knowledge.	Pupils learn best by practical lessons.	My colleagues and I regularly share ideas and materials related to science teaching.	The pupils I teach are good at retaining scientific knowledge and concepts.	The pupils in my class have lots of opportunities to revisit prior scientific knowledge.	I have a good understanding of what my pupils remember in science.	I know what has been taught in science in previous year groups.	Pupils remember scientific information well from previous year groups.	Pupils are motivated in science lessons.	Pupils are excited when talking about what they know in science.	I use ICT to support the teaching of science.
Y2	4	4	5	4	4	4	2	4	4	2	4	4	2
Y3	5	2	4	5	4	2	4	4	5	2	4	4	4
Y4	4	2	5	4	2	3	2	3	2	3	4	4	4
Y5	4	4	5	4	2	1	1	3	1	1	4	4	1
Median	**4**	**3**	**5**	**4**	**3**	**2.5**	**2**	**3.5**	**3**	**2**	**4**	**4**	**3**

TABLE 23: INTERVENTION PRE-TEST

	I enjoy teaching science.	Science should be taught through a themed topic alongside other subjects.	Pupils should be taught scientific knowledge.	Pupils learn best by practical lessons.	My colleagues and I regularly share ideas and materials related to science teaching.	The pupils I teach are good at retaining scientific knowledge and concepts.	The pupils in my class have lots of opportunities to revisit prior scientific knowledge.	I have a good understanding of what my pupils remember in science.	I know what has been taught in science in previous year groups.	Pupils remember scientific information well from previous year groups.	Pupils are motivated in science lessons.	Pupils are excited when talking about what they know in science.	I use ICT to support the teaching of science.
Y2	4	2	5	3	4	4	4	5	5	2	5	5	4
Y3	4	2	5	4	2	4	4	4	4	4	5	5	5
Y4	4	4	5	5	4	4	4	5	5	4	5	5	2
Y5	4	2	5	4	2	4	5	5	4	2	4	5	4
Median	**4**	**2**	**5**	**4**	**3**	**4**	**4**	**5**	**4.5**	**3**	**5**	**5**	**4**

TABLE 24: CONTROL POST-TEST

	I enjoy teaching science.	Science should be taught through a themed topic alongside other subjects.	Pupils should be taught scientific knowledge.	Pupils learn best by practical lessons.	My colleagues and I regularly share ideas and materials related to science teaching.	The pupils I teach are good at retaining scientific knowledge and concepts.	The pupils in my class have lots of opportunities to revisit prior scientific knowledge.	I have a good understanding of what my pupils remember in science.	I know what has been taught in science in previous year groups.	Pupils remember scientific information well from previous year groups.	Pupils are motivated in science lessons.	Pupils are excited when talking about what they know in science.	I use ICT to support the teaching of science.
Y2	4	2	5	4	4	2	2	4	4	2	4	4	4
Y3	5	3	5	5	5	2	2	4	5	2	5	4	4
Y4	4	2	5	5	3	3	2	4	4	4	5	4	2
Y5	4	5	5	5	2	2	1	2	2	1	5	5	1
Median	**4**	**2.5**	**5**	**5**	**3.5**	**2**	**2**	**4**	**4**	**2**	**5**	**4**	**3**

TABLE 25: INTERVENTION POST-TEST

	I enjoy teaching science.	Science should be taught through a themed topic alongside other subjects.	Pupils should be taught scientific knowledge.	Pupils learn best by practical lessons.	My colleagues and I regularly share ideas and materials related to science teaching.	The pupils I teach are good at retaining scientific knowledge and concepts.	The pupils in my class have lots of opportunities to revisit prior scientific knowledge.	I have a good understanding of what my pupils remember in science.	I know what has been taught in science in previous year groups.	Pupils remember scientific information well from previous year groups.	Pupils are motivated in science lessons.	Pupils are excited when talking about what they know in science.	I use ICT to support the teaching of science.
Y2	5	1	5	4	4	5	5	5	5	4	5	5	5
Y3	5	4	5	4	2	4	4	4	4	4	5	5	5
Y4	4	4	5	5	4	4	5	4	5	4	5	4	4
Y5	5	2	5	4	4	4	5	4	4	4	5	4	4
Median	**5**	**4**	**5**	**4**	**4**	**4**	**5**	**4**	**4**	**4**	**5**	**4**	**4**

While the surveys offer only a simplistic ‘snapshot’ of teacher opinions, there were a cluster of responses that were of interest when comparing between control and intervention group teachers, mainly around statements 6, 7, 8 and 10:

- The pupils I teach are good at retaining scientific knowledge and concepts.
- The pupils in my class have lots of opportunities to revisit prior scientific knowledge.
- I have a good understanding of what my pupils remember in science.
- Pupils remember scientific information well from previous year groups.

Data shows that generally intervention group teachers responded more positively to these questions compared to control group teachers. This may or not be as a result of the innovation or school culture and pedagogy.

During the observations of lessons, teachers in both intervention and control conditions delivered the first 10 minutes as planned. During control lessons, pupils worked silently and had access to both science textbooks and their school science books for re-reading. Teachers did not quiz or ask questions. Observations were similar in intervention groups although it was noted that pupils, particularly in Years 2 and 3, would often quietly cheer and ‘fist-pump’ when they answered correctly on the Socrative quiz. In contrast, feedback from control group teachers said that pupils who re-read found it ‘challenging’ to spend so long re-reading materials; that ‘some pupils found it difficult to concentrate by themselves for so long’ or that ‘pupils found it a little boring’.

Discussion

Interpretation of findings

The main purpose was to find out if the intervention would help pupils to better remember what they have been taught over time. The outcomes of the innovation are certainly promising. Three of the four intervention groups demonstrated an excellent retention after 12 weeks, with the Year 4 intervention group being hampered by technical computing difficulties. Perhaps more interesting was that disadvantaged pupils in the intervention groups were able to remember what they had been taught after a 12-week period as well as their peers. With the impact of pupil premium spending firmly in the spotlight, an intervention that can be effective is certainly encouraging.

Limitations

With a small study such as this, it would be unwise to form a generalised judgement about the impact of the intervention. However, a larger up-scaling of the innovation would produce more reliable data for evaluation. The study took place across two schools, each with their own teaching pedagogy ideologies and culture. Whereas we could control some elements of classroom climate during the intervention, it is impossible to control teacher input for each session as well as the quality of teaching and content in other subjects. For example, one question that we may consider is how much of an impact wider reading has, both in other subjects and outside of school, on intervention results.

As reported, the Year 4 intervention group was unable to access suitable technology in order to use Socrative online for four sessions which most likely would have had an impact on outcomes for the group. Furthermore, the unusual Year 5 control group data impacted on effect size for that year group and as a consequence, the effect size overall.

Finally, although published tests were considered for the outcome measure, it was decided that they were not suitable for several reasons:

- Some questions were ambiguous in nature.
- They did not fully cover the science unit taught.
- There were too few application questions.

With regards to the final point in particular, it was important to have enough application and skills questions in each test to ensure that question types did not favour the intervention group (for example, only questions on recall (concepts, facts and definitions), which would have been the case from the published examples). Consequently, tests were produced by the evaluator that included a range of question types similar to, and including, past KS1 and KS2 Standard Assessment Test science questions. In order to reduce bias, it is preferable to use standardised tests as outcome measures, and a larger study might be able to achieve this.

Implications for practice

With Amanda Spielman, Ofsted's chief inspector of education, declaring that the "vast, accumulated wealth of human knowledge, and what we choose to pass on to the next generation through teaching in our schools (the curriculum), must be at the heart of education", the highlighting of retention of learning throughout the curriculum, is clearly important. This paper would like teachers to consider:

- What key knowledge do teachers want pupils to remember?
- Are pupils given opportunities to revisit content at a later date to aid retention?
- Are systems in place to check how much children have remembered at various points in the year for subjects other than maths and English?

Teachers in schools may perceive their pupils to have learnt what they wanted them to, however, this may only be at performance level (the point of teaching or end of the lesson). Unlike maths and English in primary schools, subjects such as science may not be taught with content revisited every day. It may also be the case that foundation subject content such as history and geography does not get revisited again in the school year or at all in school (such as the Year 3 science unit, Rocks and Soils). Therefore, it is recommended that teachers provide pupils with regular, low-stakes opportunities to retrieve past content, so that it is not forgotten.

Implications for further evaluation

These positive findings raise further questions and directions of future interest. The obvious next step would be to secure funding for a larger scale project that will help to reduce variation and produce outcomes for a much larger sample size. It may be interesting to explore the effects of a variety of different retrieval practice methods. For example, if multiple-choice quizzing has a greater effect on outcomes compared to paper-based retrieval, re-reading, or the continued study of new content. It may also be of interest to explore effects with even younger children, such as Reception pupils (age 4–5) and Year 1 pupils (age 5–6), as the majority of past retrieval studies involved secondary pupils. Lastly, examining the effects of retrieval practice in other subjects would be of further interest considering many schools are re-examining their curriculum at this time.

References

Agarwal P.K, Bain P.M Chamberlain R.W (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. *Educational Psychology Review*, 24, 437–448.

Agarwal, PK, Finley JR, Rose NS and Roediger HL III (2017). Benefits from retrieval practice are greater for students with lower working memory capacity, *Memory*, 25:6, 764–771.

Benjamin AS, Bjork RA Schwartz BL (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, 127, 55–68.

Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ Willingham DT (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14, 4–58.

Jacoby LL (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, 17, 649–667.

Karpicke, JD, & Roediger, HL (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention.

McDaniel MA., Agarwal PK., HuelserBJ, McDermott KB, Roediger HL, III (2011). *Journal of Educational Psychology*, 103:2, 399–414.

Roediger et al. (2011).Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology*: 382–395.

Roediger HL and Karpicke JD (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science* 17: 249–255.

Whitten WB Bjork RA (1977). Learning from tests: Effects of spacing*. Journal of Verbal Learning and Verbal Behavior*, 16, 465– 478.

Appendices

Appendix 1: Intervention timeline

Intervention group			Control group	
Week 1	Pupil and teacher questionnaire before baseline science test A		Week 1	Pupil and teacher questionnaire before baseline science test A
Week 2 Observations	Begin teaching science unit, no retrieval practice.		Week 2	Begin teaching science unit, no re-reading.
Week 3	5 mins retrieval practice via Socrative. Continue teaching of new content in unit.		Week 3	5 mins re-reading. Continue teaching of new content in unit.
Week 4	10 mins retrieval practice via Socrative. Continue teaching of new content in unit.		Week 4	10 mins re-reading. Continue teaching of new content in unit.
Week 5	10 mins retrieval practice via Socrative. Continue teaching of new content in unit.		Week 5	10 mins re-reading. Continue teaching of new content in unit.
Week 6	10 mins retrieval practice via Socrative. Continue teaching of new content in unit. Pupil questionnaire.		Week 6	10 mins re-reading. Continue teaching of new content in unit. Pupil questionnaire.
Week 7	10 mins retrieval practice via Socrative. Continue teaching of new content in unit.		Week 7	10 mins re-reading. Continue teaching of new content in unit.
Week 8	10 mins retrieval practice via Socrative. Continue teaching of new content in unit.		Week 8	10 mins re-reading. Continue teaching of new content in unit.
Week 9	End of unit test B		Week 9	End of unit test B
Week 20	Pupil questionnaire. Repeat end of unit test A Teacher questionnaire		Week 20	Pupil questionnaire. Repeat end of unit test A. Teacher questionnaire

Appendix 2: Pupil survey

Do you enjoy science?

Are you good at science?

Do you know lots of stuff about science?

Can you remember lots of science stuff from last year?

Are you good at answering science questions?

How do tests make you feel?

Appendix 3: Teacher survey

	Strongly Disagree	Disagree	No opinion	Agree	Strongly Agree
I enjoy teaching science.					
Science should be taught through a themed topic alongside other subjects.					
Pupils should be taught scientific knowledge.					
Pupils learn science best by practical lessons.					
My colleagues and I regularly share ideas and materials related to science teaching.					
The pupils I teach are good at retaining scientific knowledge and concepts.					
The pupils in my class have lots of opportunities to revisit prior scientific knowledge.					
I have a good understanding of what my pupils remember in science.					
I know what has been taught in science in previous year groups.					
Pupils remember scientific information well from previous year groups.					
Pupils are motivated in science lessons.					
Pupils are excited when talking about what they know in science.					
I use ICT (computers/iPads) to support the teaching of science lessons.					

Contact us

+44 (0)1904 328166 info@the-iee.org.uk
Berrick Saul Building, University of York, York YO10 5DD
Twitter: @IEE_York
website: the-iee.org.uk

The Institute for Effective Education (IEE) is an independent charity working to improve education for all children by promoting the use of evidence in education policy and practice.

In collaboration with the Education Endowment Foundation (EEF) we support a national Research Schools Network and have developed resources aimed at people on the front line of education.

The Institute for Effective Education is a charity registered in England, charity number 1168744

Institute for
Effective Education
Empowering educators with evidence

Research
Schools Network