



Institute for
Effective Education
Empowering educators with evidence

Low stakes quizzing

To what extent does low stakes quizzing in Key Stage 3 geography and history affect meaningful learning?

Mark Enser and Emma Smith
Heathfield Community College
November 2020



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

<i>Section</i>	<i>Page</i>
Executive summary	4
Description of the innovation	4
Summary of the evaluation	4
Summary of findings	4
Costs	5
Introduction	6
Description of the problem	6
Review of existing research	6
Description of the innovation	6
Research question	7
Method	8
Sample	8
Allocation to groups	8
Innovation	8
Timeline	9
Outcome measures	9
Process evaluation	10
Analysis	10
Costs	11
Results	12
Process evaluation findings	14
Discussion/Conclusion	15
Limitations	16
Implications for practice	16
Implications for further evaluation	17
Conclusions	17
References	18

Executive summary

Description of the innovation

The evaluation sought to establish whether weekly low-stakes quizzing of Year 8 geography and history pupils across five half-terms would lead to an improvement not only in the breadth of knowledge but also its application during end-of-year assessments in both subjects. The evaluation also sought to establish whether this form of quizzing is sustainable in everyday classroom practice.

Summary of the evaluation

Heathfield Community College is a rural, comprehensive secondary school in the south of England. The Year 8 pupils taking part in the study were a mixed prior-attaining cohort and were taught in eight mixed prior-attainment classes throughout the academic year in geography and history by subject specialists. Four classes (approximately 50% of the cohort) were randomly selected to receive the weekly, 10-minute low-stakes quizzing intervention which consisted of 10 closed questions. The remaining four classes were the control classes who only took part in retrieval practice on an ad hoc basis throughout the evaluation period. All pupils sat a baseline assessment in each subject at the start of the evaluation and these mean baseline scores were compared with the mean scores achieved in the end of year assessments in each subject.

Summary of findings

The intervention in geography showed that it did have an impact with an effect size of +0.3 for all pupils. It had a greater effect on girls (+0.4) and on pupils who are not regarded as pupil premium (+0.5). The greatest effect (+0.8) was on pupils with special educational needs (SEND), but it has to be considered that the sample size was small ($n=9$), especially compared with the whole cohort ($n=218$). Our key conclusion is that low-stakes retrieval practice in the form of short-answer recall tests do boost meaningful learning in geography. My view is that this is because they enable pupils to use what they know to better understand and write about the subject.

In contrast, the intervention in history showed that it had a much smaller impact with an effect size of +0.1 for all pupils. It had a slightly greater impact for boys (+0.2) in comparison to girls (+0.1). Similar to geography, the greatest effect was on pupils with SEND (+0.2), but the sample size was also small ($n=12$) in comparison to the whole cohort ($n=215$). Consequently, the conclusion drawn here is that low-stakes retrieval practice in the form of short recall tests in history leads to a smaller increase in meaningful learning and may be of less help with pupils applying knowledge to write about the subject than in geography.

One limitation of the study is that the conclusions are predicated on the final assessment truly being a test of meaningful learning and this could be contested. There is also the need to be very clear about what we mean by “works” – works for what purpose? (Biesta 2015). In the case of geography, I would say that our study shows that regular quizzing of previously studied material led to pupils being able to write better answers to geographical questions.

Although the overall sample sizes for the intervention group and the control group were fairly large compared to other practitioner research studies some of the subgroups were much smaller, such as the numbers of SEND and pupil premium pupils. There were differences in the numbers of pupils who eventually sat the history and geography assessments due to pupil absence. This is particularly seen where the number of SEND pupils who successfully completed all parts of the study was different in geography ($n=9$) to history ($n=12$). Similarly the final number who sat the history end-of-year assessment was lower ($n=215$) than those

who sat the geography end-of-year assessment (n= 218), as not all pupils in the cohort completed the end-of-year assessment.

Costs

The costs of this intervention are low. There is a time cost in creating low-stakes quizzes but this need be no more time-consuming than planning any other activity. Time is also needed to explain the purpose of the retrieval quizzes to all members of staff to ensure that they are carried out with the right intention in mind (low stakes and with the purpose of encouraging retrieval rather than as a form of formative assessment).

Our intervention had increased costs due to the completion of tests on paper for ease of checking that they were being completed as per instruction but they could have just as easily have been done in the back of exercise books or digitally using Google Forms.

Introduction

Description of the problem

Pupil knowledge retention from one lesson to the next, from topic to topic and by the end of each academic year varies, having implications for pupil progress throughout the academic year. Some pupils struggle to recall knowledge acquired from a previous topic and then apply it when developing knowledge and skills in a later one. Consequently, pupils can end Years 7 and 8 with only surface knowledge of history and geography which limits their progress and has longer-term implications for further study at Key Stage 4.

Review of existing research

There is a significant amount of research about retrieval practice. Rawson and Dunlosky (2012) recommend that pupils should practise target knowledge until it is correctly recalled once. This would suggest that pupils in the classroom should have the opportunity to regularly practise retrieving knowledge until it is learned, requiring more than one opportunity to retrieve it in a low-stakes quiz, over a period of time. Despite taking up lesson time, Daniel and Broida (2004) found that in-class quizzing for 15 minutes a week over a 16-week period had a positive impact on pupil outcomes. Roediger & Karpicke (2006) found that retrieval practice was more effective than time spent studying on delayed testing performance (for tests taken two days or one week later). Roediger et al (2011) identified ten benefits that accrue from testing that go beyond improved recall and include increased pupil awareness of gaps in their knowledge and organisation of ideas.

At GCSE history level, Donaghy (2014) also found that using regular, low-stakes testing with one Year 10 class in four-week cycles improved their knowledge retention, how they applied knowledge in summative assessment and pupils' progress in contrast to a control group. Dennis (2016) regularly used low-stakes testing with two Year 8 history classes for the first term of an academic year, which appeared to cause improvement in pupil retention and progress in securing knowledge. But pupils who had been regularly tested in the first term could not always apply the knowledge effectively in an assessment six months later in a question that was different to what they had previously encountered, performing no better than the control group.

Description of the innovation

Consequently, the available research suggests that regular low-stakes quizzing has a positive impact on pupil progress, but its impact on developing a conceptual understanding of the subject is unclear. Instead, this innovation would use low-stakes quizzing regularly to recall core knowledge over an academic year with a large cohort of Key Stage 3 pupils in two humanities subjects simultaneously with the aim of establishing whether or not this testing effect leads to an improvement not only in the breadth of knowledge but also its application. Simultaneously, it would also test if this form of quizzing is sustainable in everyday classroom practice.

Our innovation involved Year 8 pupils taking a short, 10-question, open-answer quiz at the start of their lessons once a week in geography and history. The questions were given to them on a piece of paper with room for them to record a short response. The answers were then displayed and pupils corrected anything they had got wrong. These questions were drawn from previous lessons and, as the innovation went on, previous topics. These quizzes were kept low stakes and pupils were not expected to reveal their final score.

Research question

The innovation sought an answer to the question: what impact does low-stakes quizzing for 10 minutes a week over four half terms, compared to less rigid, ad hoc retrieval practice, have on the application of subject knowledge and progress in Year 8 history and geography pupils?

Our primary question is:

- Does regular retrieval practice through low-stakes quizzes lead to meaningful learning in geography and history (defined as “If a person has learned something, it means they are capable of using information available in a particular context, referred to as retrieval cues, to reconstruct knowledge in order to meet the demands of the present activity” (Karpicke and Grimaldi, 2012))?

Our secondary questions are:

- Does retrieval practice have the same impact despite gender, economic background, special educational needs (SEND) or prior attainment?
- Are low-stakes quizzes a manageable form of intervention in Key Stage 3 history and geography lessons?

Method

Sample

Heathfield Community College is an 11–18 mixed comprehensive school in the rural town of Heathfield, East Sussex in the south of England. In 2018/19, the school had:

Total number of pupils on roll (all ages)	1,439
Girls on roll	52%
Boys on roll	48%
Pupils with a SEN education, health and care plan	2.3%
Pupils with SEN support	4.2%
Pupils whose first language is not English	1%
Pupils eligible for free school meals	6.2%

Heathfield Community College performs above the national average at both GCSE and A-level.

The Year 8 cohort at Heathfield Community College (HCC) in 2018/19 academic year consisted of 229 pupils aged 12–13 years. The cohort was 51% male, 49% female. Of the cohort 5.2% had a special educational need or difficulty (SEND) and 16.1% received pupil premium funding. Pupils were arranged into the same mixed-prior-attainment classes for history and geography.

Allocation to groups

To allocate the innovation and control groups, classes were selected randomly. The same four classes were given the innovation and the same four classes were in the control group in both geography and history. Allocation to groups was completed once timetables had been finalised in the academic year so the assignment of teachers to classes was already in place.

Innovation

Year 8 pupils were taught three history lessons and three geography lessons every fortnight. Our innovation involved all pupils in four intervention classes starting a lesson once a week with a 10-question quiz. The questions on these quizzes were drawn from previous lessons and previous topics (with five questions from within the current topic and five from previous topics) and were set by the head of department. The questions were based on the recall of key names, dates, people, places and statistics or short procedural questions such as “Why is it hotter on the equator?” or “How do stalactites form?” The quiz was printed so that pupils could write their answers next to the questions. At the end of the quiz, the answers were revealed and pupils marked their own work correcting any answers that were wrong. The papers were then collected in, primarily so that we could monitor fidelity to the innovation. Pupils were frequently reminded that the purpose of these quizzes was not to assess them and that we weren’t collecting their scores. We explained that the purpose was to see if it helped them to remember things they had studied for longer.

The heads of department also had a role in training teachers in the use of the quizzes and ensuring fidelity to the innovation with termly checks on the quizzes. They also taught classes in Year 8.

This innovation ran from the beginning of term 2 until the end of term 5. At this point all pupils in both the control and innovation group sat an assessment based on everything they had studied that year. In geography these topics were: portrayals of East Africa, creation of landscapes, and

Russia power and borders. In history they studied: the campaign for women’s suffrage in Britain the 19th / 20th centuries, the causes and events of the First World War, the Holocaust, the significance of key battles in the Second World War, and the USA in the 1920s.

The intervention classes were taught by three teachers in geography and two in history. These teachers and those teaching the control groups had an hour’s training before the innovation began on the purpose of retrieval practice and on the protocol for the innovation. We made it clear that only the innovation group should use the quizzes and that everything else should be taught as normal. Quizzes were written and copied for the teachers carrying out the innovation by the heads of department.

Timeline

Date	Activity
Term one (September–October 2018)	Pupils study their first topic
Final week of term one (late October 2018)	Baseline test on this topic
Terms 2–5 (November 2018– April 2019)	Intervention group classes have weekly low stakes quizzes
Final week of term 5 (late May 2019)	All pupils sit end of year assessments in history and geography

Outcome measures

All participating pupils took pre-tests in the final week of term 1 and post-tests in the final week of term 5. At both points, pupils took both history and geography tests. The pre-tests and the post-tests in both subjects were out of 40 marks and the questions sat were GCSE in format ie. with a mix of questions of different tariffs designed to test different aspects of the subject. Both were sat in the same conditions and marked in the same way (see below).

These tests were written by the heads of department for the two subjects at the end of the preceding year and were then given to the exams officer. The papers weren’t released to the other teachers in the department or available for reference by the heads of department.

The pre-test comprised topics studied over the course of term 1. Following the pre-test pupils were provided with feedback and made corrections according to the assessment policies of the geography and history departments.

The effect of the innovation was measured using an end-of-year assessment in both geography and history. All assessments were closed-book exams, sat in the hall in exam conditions. These assessments were made up of a range of questions from the first term in the pre-test and all topics studied that year in the post-test and both were designed to test not just recall of information that had been quizzed in the innovation but its wider application to more in-depth questions. For example, the geography post-test included the question “Why did Russia annex Crimea?” which required them to draw on a range of different ideas from the topic (on both physical and human geography) and draw it together to reach a conclusion.

These questions were marked in the manner of GCSE exam questions with levelled scores depending on the complexity of the answer and with a total score of 40 for both assessments. The assessments were marked blind by all members of both departments with no one knowing which were from the control and which were from the innovation group. To ensure correct

application of the mark scheme for both assessments, standardisation meetings were held and the head of department checked a selection of the marking.

Process evaluation

The innovation was delivered as intended. This was monitored through the collecting in of completed quizzes and through discussions with pupils. Three intervention history classes had lessons on either Monday or Friday which, due to term-time dates, bank holidays and INSET days meant that not all quizzes were completed by these three groups. One group missed one of the 30 quizzes, while the two other classes missed four. The non-intervention history classes were not affected in this way, having their classes mid-week.

Follow-up sessions were used throughout the year to check fidelity to the innovation plan and checks were made that the quizzes were being completed correctly where possible. These involved the heads of department looking at completed pupils' quizzes from all classes at end of terms 2, 3, 4 and 5, although no feedback was given to the teachers delivering the intervention. Lesson drop-ins were also completed by the heads of department in terms 2 and 4 to check that the intervention was not being carried out in control classes.

The teachers' views of completing the quizzes were evaluated through the use of an anonymous survey completed by those who carried out the innovation. This was five teachers in total: three geography teachers and two history teachers. Teachers were asked to rate the following statements using a Likert scale:

1. Completing the weekly quizzes was straightforward to manage as a teacher.
2. My pupils completed all of the weekly quizzes provided.
3. Completing the weekly quizzes took up a lot of lesson time.
4. The completion of the weekly quizzes disrupted the teaching of topics
5. The completion of the quizzes helped to secure pupils' knowledge across the academic year
6. The completion of the quizzes helped pupils to approach the end-of-year assessment with confidence

Analysis

The analysis focused on comparing the mean results of the control group and the innovation group on the final assessment. This was used to generate an effect size based on average progress made between the pre-test and post-test. We did this for the whole cohort and then for particular groups to allow us to compare impact depending on gender, pupil premium status, SEND and prior attainment (prior attainment was based on Key Stage 2 data where available).

Costs

The costs of the evaluation include:

Budget Item	Amount
Twilight CPD for all teachers involved in the innovation 1hr x 9 geography/ history teachers. Average teacher salary.	Approx. £400
Creation of weekly quizzes by heads of department throughout the academic year. Six hours for each head of department. Average teacher salary.	Approx. £1,000
Moderation and standardisation of assessments (2hrs). 9 history/ geography teachers, average teacher salary.	Approx. £800
Resources: including photocopying	Approx. £1,000
Total approximate expenditure	£ 3,200

Results

The mean test scores of the pupils' pre-test and end-of-year tests for geography and history are detailed in the tables below (all scores shown to one decimal place).

Only pupils who completed both the pre-test and post-test had their data included in the data analysis: 218 pupils completed both geography tests and 215 pupils completed both history tests. Prior attainment data for two history pupils and 20 geography pupils was unknown: these pupils were included in all data analyses except the subgroup analyses by prior attainment.

It is worth keeping in mind that the pre-tests and post-tests were not equivalent in terms of difficulty. The pre-test only covered one topic whereas the post-test covered material from a greater range of topics and involved pupils thinking across topics.

FIGURE 1: GEOGRAPHY TEST RESULTS

		Mean pre-test score / 40	Mean post-test score / 40	Mean progress
Whole cohort	Intervention group (n = 110)	20.3	17	-3.3
	Control group (n = 108)	21	15.7	-5.3
Boys	Intervention group (n = 54)	19.3	16.8	-2.6
	Control group (n = 53)	18.2	13.9	-4.3
Girls	Intervention group (n = 56)	21.3	17.1	-4.2
	Control group (n = 55)	22.6	16.2	-6.4
Pupil premium	Intervention group (n = 21)	19.9	16.4	-3.5
	Control group (n = 19)	20.1	15.5	-4.6
Non-pupil premium	Intervention group (n = 89)	20.4	17.8	-2.6
	Control group (n = 89)	20.9	15.7	-5.2
SEND	Intervention group (n = 5)	10.6	9.2	-1.4
	Control group (n = 4)	8.3	3.5	-4.8
Non-SEND	Intervention group (n = 105)	20.6	17.1	-3.5
	Control group (n = 104)	21.6	16.2	-5.4
Higher prior-attaining	Intervention group (n = 40)	24.4	21.4	-3
	Control group (n = 33)	25.4	21.8	-3.6
Middle prior-attaining	Intervention group (n = 30)	20.9	16.6	-4.3
	Control group (n = 40)	21.6	16	-5.6
Lower prior-attaining	Intervention group (n = 29)	15.7	12.6	-3.1
	Control group (n = 26)	15.2	11.8	-3.4

FIGURE 2: HISTORY TEST RESULTS

		Mean pre-test score / 40	Mean post-test score / 40	Mean progress
Whole cohort	Intervention group (n = 108)	16.3	16.2	- 0.1
	Control group (n = 107)	17.7	16.9	-0.8
Boys	Intervention group (n = 54)	15.0	15.4	0.4
	Control group (n =53)	15.4	14.7	- 0.7
Girls	Intervention group (n = 54)	17.5	17.1	- 0.5
	Control group (n = 54)	20.0	19	-1.4
Pupil premium	Intervention group (n = 14)	15.0	13.9	-1.1
	Control group (n = 19)	13.2	12.8	- 0.4
Non-pupil premium	Intervention group (n =94)	16.5	16.6	0.1
	Control group (n =88)	18.7	17.7	-0.1
SEND	Intervention group (n =3)	8.3	8.6	0.3
	Control group (n = 9)	10.8	10.2	-0.7
Non-SEND	Intervention group (n = 105)	16.5	16.4	-0.06
	Control group (n =98)	18.3	17.5	-0.9
Higher prior-attaining	Intervention group (n = 41)	19.9	20.1	0.3
	Control group (n = 34)	22.3	22.4	0.1
Middle prior-attaining	Intervention group (n = 35)	16.2	15.7	-0.5
	Control group (n = 45)	17.4	16.1	-1.4
Lower prior-attaining	Intervention group (n =31)	11.6	11.4	-0.2
	Control group (n = 27)	12.5	11.6	-0.9

FIGURE 3: EFFECT SIZES

	Geography effect size	History effect size
Cohort	0.3	0.1
Boys	0.3	0.1
Girls	0.4	0.1
Pupil premium	0.3	-0.1
Non-pupil premium	0.5	0.1
SEND	0.8	0.2
Non-SEND	0.3	0.1
Higher prior-attaining	0.1	0
Middle prior-attaining	0.2	0.1
Lower prior-attaining	0.1	0.1

Process evaluation findings

All intervention group teachers (three geography and two history teachers) completed the anonymous questionnaire. Their responses to the questionnaire are as follows:

	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
Completing the weekly quizzes was straight forward to manage as a teacher.	40%	60%	0	0	0
My pupils completed all of the weekly quizzes provided.	60%	40%	0	0	0
Completing the weekly quizzes took up a lot of lesson time.	0	20%	20%	60%	0
The completion of the weekly quizzes disrupted the teaching of topics.	0	0	40%	40%	20%
The completion of the quizzes helped to secure pupils' knowledge across the academic year.	20%	40%	40%	0	0
The completion of the quizzes helped pupils to approach the end of year assessment with confidence.	0	40%	60%	0	0

Teachers were also asked to state at the start of the questionnaire whether they had taught history or geography as part of the innovation. The history teachers fed back in the questionnaire that not all of the quizzes that were set were completed. This is unsurprising with three of the classes, two of which had history every Friday, and the other class had history on a Monday every other week. School holidays (including bank holidays) and INSET had a disruptive effect on these classes completing not only the innovation as initially planned but also the delivery of the planned Year 8 curriculum. Teachers in both history and geography were less confident that the regular quizzing helped secure pupils' knowledge or helped their pupils to approach the end-of-year assessment confidently. This could be because they were aware that they were taking part in a project to test the efficacy of these tests and so are aware it is contested. However, all teachers taking part in the innovation had a positive outlook on the manageability of delivering the quizzes on a regular basis in lessons.

Discussion/Conclusion

The findings in geography seem to support the research literature that retrieval practice does lead to meaningful learning. This isn't surprising as geography often concerns itself with applying what we know about the world to new situations. In this instance, pupils were regularly quizzed on things such as Russia's location, its climate and the barriers its geography created for development. This led to more in-depth answers to such assessed question as "Why did Russia invade Crimea in 2014?", with the mark scheme giving more marks for answers showing greater depth.

All groups in geography showed a positive effect size but it seemed to have a particularly marked impact on girls and pupils with special educational needs (SEND). The increased effect size on pupils with SEND has to be taken with the reminder that the sample size was small ($n=9$) but the results do seem to support some of the research literature that suggests pupils who struggle with cognition may benefit from increased support for working memory (see for example my discussion, Enser, 2019). One way to achieve this support is to develop stronger links between topics (to develop their schema) and to make retrieval from this schema easier through regular practice.

The explanation for the increased effect on girls is harder to find and contrasts with the results in history. It is conceivable that the presence of the retrieval quizzes prompted a higher proportion of female pupils to start revising their notes so as to be prepared for them or that they took the quizzes more seriously having realised their impact. This would be supported by research from Griffin et al (2012) and Gagnon and Cormier (2018) which found that girls were more likely to use distributed practice throughout the course.

The history results are less conclusive. An effect size of +0.1 is a modestly positive effect size, showing that over the course of the evaluation the intervention group pupils made slightly more progress in the application of historical knowledge than control group pupils. We would propose that this modest impact of regular quizzing in history in comparison to geography could be explained by a combination of factors: the number of history quizzes not completed due to school holidays and/ or INSET days, less experienced teachers delivering the innovation in comparison to the teachers teaching the control classes and the breadth of the Year 8 history curriculum compared to that of geography.

However the outcomes in history could support Dennis' (2016) findings that regular low-stakes testing cannot solely help pupils transfer knowledge and apply it meaningfully, although Dennis' was a smaller study. In this evaluation it would seem that this was especially the case in regards to questions that demand source analysis and those that required pupils to use the knowledge supposedly secured through the weekly quizzes, applying it to a broader context. So, although the intervention quizzes regularly asked *What shape were trenches?* (on four knowledge quizzes) and *What shape were the trenches ?* (six knowledge quizzes), these features were rarely found in the pupil responses to the straightforward question *Describe the fighting in the trenches in the First World War*. Instead, many responses focused on the muddy conditions and trench foot which, although in the innovation *trench foot* appeared in four weekly quizzes, this was not the most appropriate or sophisticated knowledge to apply to address the end-of-year assessment question. Only 40 out of 108 responses from the innovation group included knowledge beyond trench foot or more generic descriptions of the weather/muddy conditions or falling bombs. It would be reasonable to suggest that the characteristics of mud and trench foot had more easily fit into the pupils' developing schema of the First World War when pupils were first taught the topic (in term 2) and that these then were more readily recalled

when pupils saw *First World War* in the exam question. It could also be suggested that in history the quizzes, which had to cover core knowledge from five different topics compared to three from geography, had not actually helped to locate the World War One facts in pupils' existing schema and so pupils were not able to easily recall them and apply them suitably in the end-of-year assessment. Another parallel with Dennis' study is that the majority of pupils in both intervention and control classes alike were most confident in answering the question on women's suffrage on the end-of-year assessment; a topic that had been the sole focus of the baseline assessment in term 1. So the quiz questions on this topic had potentially supported pupils to recall and then meaningfully answer an assessment question on a topic that they had previously encountered. This is perhaps unsurprising as, following the pre-test, pupils were provided with feedback and corrected their responses.

Limitations

One limitation was the fairly small sample size, especially for sub-groups such as the low number of pupil premium pupils and those who were lower prior-attaining.

There was also missing data for pupils who missed the post-test at the end of the year meaning these pupils were not included in the analysis as well as pupils joining the cohort who would not have completed the pre-test. It is conceivable that these pupils are not representative of the whole sample and therefore their missing data have influenced the results.

The history innovation classes were taught by different teachers than the control classes; although one teacher who taught three of the innovation classes also had taught one control class. On average, there were more experienced teachers teaching the history control groups than then the history intervention groups. It is possible then that differences in general teacher effectiveness led to the minimal effect of the innovation in history, particularly regarding ad hoc teacher questioning of more experienced teachers of the control classes where retrieval practice may have occurred as part of regular practice. However, contrary to this statement, there was some disruption to one control class due to a teacher absence, although the class teacher was replaced by another, experienced history specialist. In addition, the disruption to the completion of the quizzes and missing taught hours due to non-school days is also likely to have had an impact.

In geography, two of the teachers who had innovation classes also had control classes. It is possible that they increased the use of retrieval practices with the control classes in other forms than the regular quizzing and so led to an enhanced effect for the control. If this were the case then the innovation had an even greater impact than our results suggest.

The two tests were written by the same people who wrote the intervention quizzes. It is possible that this led to a subconscious tendency to include questions in the quizzes that echoed the content of the post-test.

The analysis doesn't take into account pupil attendance. This would only be significant if attendance was lower for certain groups of pupils who regularly missed the quizzes. It might mean that the effect of the innovation would have been higher for these groups had they attended. However, this would also highlight a weakness of this particular type of innovation for these groups.

Implications for practice

The findings from this innovation in geography will lead to us continuing with low stakes quizzes throughout Key Stage 3 and increasing their regularity in Key Stage 4 and Key Stage 5. The difference between the history and geography findings would suggest that other subjects should trial and evaluate the use in their own subjects before adopting them as a part of regular

practice. These findings suggest that the research literature on the importance of retrieval practice in meaningful learning (Karpicke and Grimaldi, 2012) applies to geography and it will lead to us continuing to explore how it can be embedded not just in the form of these quizzes but across our curriculum. The findings also suggest that low-stakes quizzing might not be sufficient on its own in history to help pupils apply knowledge competently in summative assessments. It could be beneficial in history for low-stakes quizzing to be immediately followed up by asking broader questions that demand pupils to apply knowledge, rather than always quizzing knowledge in isolation. This would require careful curriculum planning and the development of subject-specific pedagogy. However, more research should be carried out into the potential impact of this approach. Regardless of the subject, regular low-stakes quizzing is easy to implement in the classroom and so should not be avoided on the grounds of classroom management.

Implications for further evaluation

Our study suggests that starting a lesson with retrieval practice through low-stakes quizzing may be more beneficial in some subjects than others. It would be useful to widen this study out to consider other subjects and whether these quizzes would be beneficial there. It is entirely possible that these quizzes are also beneficial in history and further research in this subject is needed in which variations in teacher quality and the number of lessons delivered is controlled.

This study indicated that these quizzes had a large effect on pupils with SEND but the sample size was small. Further research would be needed to see if this was replicated in a larger sample and to determine which particular needs this form of retrieval quizzing was supporting.

Conclusions

We conclude that in geography low-stakes retrieval quizzing leads to not only better retention of information but also an improved ability to apply it to new contexts. Our results for history are less clear-cut; more research is needed to understand the potential impact of low-stakes quizzing in history and whether variations of a retrieval practice approach may have greater impact on the retention and application of historical knowledge.

These low-stakes quizzes in geography led to benefits for all groups of pupils but they had a greater effect on some than others. Pupils with a SEND designation, non-pupil premium pupils and girls benefited more from the innovation than other groups and prior attainment seemed to make little difference.

References

Daniel D B & Broida J (2004). Using web-based quizzing to improve exam performance: lessons learned. *Teaching of Psychology*, 31(3), 207–208

Dennis N (2016). Cognitive psychology and low stakes testing without guarantees. *Teaching History*, 164: 22–28

Donaghy L (2014). Using regular low-stakes testing to secure pupils' contextual knowledge in Year 10. *Teaching History*, 157: 44-51

Enser M (2019). 5 Ways to support pupils with poor working memory, TES, available at <https://www.tes.com/news/5-ways-support-pupils-poor-working-memory>

Gagnon M & Cormier S (2019). Retrieval practice and distributed practice: the case of French Canadian students. *Canadian Journal of School Psychology*, 34(2), 83–97

Griffin R (2012). Do learning and study skills affect academic performance? – an empirical investigation, *Contemporary Issues In Education Research* (Vol 5 No.2) 109–116

Karpicke J D & Grimaldi, P.J (2012) Retrieval-based learning: a perspective for enhancing meaningful learning, *Educational Psychological Review*, 24:401–418

Roediger H & Karpicke J (2006). Test-enhanced learning: taking memory tests improves long term retention, *Psychological Science*, 17 (3): 249–255

Roediger H, Putnam A & Sumeracki M. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation*, 55 1–36.

Contact us

+44 (0)1904 328166 | info@the-ieee.org.uk
Ron Cooke Hub, University of York, York YO10 5GE
Twitter: @IEE_York | the-ieee.org.uk/

© Institute for Effective Education, 2020

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

