Retrieval practice in science

The North East Learning Trust

Introduction

Problem: Describe the problem or issue your innovation addresses.

What challenge(s) do your school(s) have that need to be addressed?

Ashington Academy and Bedlington Academy (both previously judged grade 4 by Oftsed) have recently been incorporated into The North East Learning Trust, a multi-academy trust that has at its centre Shotton Hall Research School. Teaching with a focus on developing memory is a core teaching and learning (T&L) priority for Ashington and Bedlington Academies. Ashington Academy has low attainment and progress in science at Key Stage 4 (P8 -0.76 in 2018). Disadvantaged pupils perform significantly less well than their non-disadvantaged peers (P8 -1.34 compared with -0.6 for non-disadvantaged). Low prior-attaining pupils perform notably less well than the rest of the cohort (P8 -1.03). While overall progress and attainment is higher in science at Bedlington Academy, there is still a significant gap between the performance of disadvantaged pupils and their non-disadvantaged peers. In both schools, staff identify pupils' lack of resilience and poor retention of core knowledge as a significant barrier to progress, particularly when faced with new, more demanding GCSEs. Previous preferred teaching and learning approaches in school have been more focused on skill development than knowledge development.

Bedlington Academy has comparatively good attainment and progress in science at Key Stage 4 (P8 0.26 in 2018). However, disadvantaged pupils perform notably less well than their non-disadvantaged peers (P8 -0.53 compared to 0.72 for non-disadvantaged).

After a cross-MAT randomised controlled trial (RCT) and investigation of the rigorous evidence base around cognitive science and memory, we have sourced best practice in terms of memory, retention and recall. As part of our teaching and learning approach and lesson planning format, all schools currently use a SMART (Smart Minds Active Recall Time) retrieval activity at both the start and end of lessons, where the starts utilise low-stakes quizzing to test previous learning, focusing on interleaving, and the ends of lessons use quizzing to test on learning within the

lesson. This is showing promising progress in terms of retention of core knowledge and vocabulary that serve as successors to future learning. We propose to investigate if Learning by Questions (LbQ) will aid greater retention and retrieval than the existing SMART retrieval tasks.

Existing evidence: What existing research evidence exists?

What does the existing research evidence say about this problem and how it could be addressed?

Learning is a complex activity which should have two core outcomes:

- the long-term retention of valuable knowledge, skills or understanding

- the ability to transfer what has been retained to different contexts and situations.

Thus, because knowledge is the platform upon which all learning rests, transfer from the short-term memory into the long-term memory is of pivotal importance. There is evidence that retrieval practice in the form of tests used as learning events, rather than just as outcome measures, can enhance pupil learning and aid the transfer of information (Karpicke and Blunt, 2011; Roediger & Karpicke 2006). To be able to retrieve, use, and apply knowledge in the long term, it is highly effective to practise retrieving, using, and applying knowledge during learning (Karpicke 2014). Furthermore, many studies e.g. Karpicke (2011) point to the potential impacts of retrieval practice over-elaborate or repeated study.

Thus, retrieval practice is of significant value in educational settings and research evidence continues to confirm that using retrieval practice improves pupil learning in different educational contexts. Several other studies have shown promising results for the impact of quizzing on retention (McDaniel et all 2012, Trumbo et al 2016), although these have largely been focused on older secondary pupils and university students. We would like to investigate whether there is a similarly positive impact on younger, less able pupils through a small-scale cross-MAT RCT.

A randomised controlled trial (RCT) conducted by Shotton Hall Research School at Bedlington Academy and Ashington Academy in October–December 2017 tested whether daily retrieval practice had a positive impact on vocabulary retention for Year 8 pupils studying English literature. The findings from this study indicated a positive effect size of +0.2 (with greater positive effect of +0.5 for high and low prior-attaining pupils). The gains from the intervention in this trial were large for higher and lower prior-attainment pupils, suggesting that retrieval practice has a positive impact on the attainment of children in these groups. Overall, the RCT results concur with the findings from previous research evidence (eg. Karpicke 2011).

In this instance, we are interested in further investigating if retrieval practice can have a positive impact on middle-attaining pupils by using LbQ, as well as further developing our work around lower-attaining pupils.

Innovation: Describe the innovation you will evaluate.

Give a brief description of the innovation. What existing research suggests the innovation will help improve the problem you have identified and benefit teachers and learners?

We aim to use LbQ for memory retrieval at the beginning of lessons, substituting for current traditional retrieval activities used (see above for research). Currently, the retrieval method used in school is effective, but only when used consistently and systematically. The instant feedback feature of LbQ will allow us to examine whether more structured, consistent feedback allows for greater retention and retrieval.

We aim to use LbQ as a starter task for Year 10 classes.

Research question(s) or hypothesis

What effect will the intervention, implemented for how long, with which pupils, have on which outcomes?

Is LbQ a more effective method of retrieval practice in science than more traditional SMART Connect activities in Year 10 pupils, as measured on attainment scores after six months?

Is LbQ a more effective method than more traditional SMART Connect activities of raising resilience of learners in Year 10 after six months?

Null hypothesis

Over six months using LbQ and its instant feedback-feature as the medium for a quiz in every lesson does not improve pupil retention with mid to lower prior-attaining pupils in science. It is less effective than traditional SMART Connect retrieval activities.

Experimental hypothesis

Over six months using LbQ and its instant feedback-feature as the medium for a quiz in every lesson does improve pupil retention with mid to lower prior-attaining pupils in science. It is more effective than traditional SMART Connect retrieval activities.

Method: Describe exactly how the evaluation will be conducted

Sample

Who will participate in your evaluation? What consent will be gathered for them to participate? Briefly describe the key characteristics of the setting and participants.

Participants

Participants will be sourced from Ashington Academy and Bedlington Academy. Ashington Academy has low attainment and progress in science at Key Stage 4 (P8 -0.76 in 2018). Disadvantaged pupils perform notably less well than their non-disadvantaged peers (P8 -1.34 compared with -0.6 for non-disadvantaged). Low prior-attaining pupils perform significantly less well than the rest of the cohort (P8 -1.03). While overall progress and attainment is higher in science at Bedlington Academy, there is still a notable gap between the performance of disadvantaged pupils and their non-disadvantaged peers.

Across both schools, four classes will be case-matched as closely as possible based on Key Stage 2 APS and pre-test data. All classes will be from Year 10, and both schools will contain both the intervention and control groups. This will be randomly allocated. All classes are setted by ability, generally by Key Stage 2 APS and all sets are potential participants based on how closely their Key Stage 2 APS align. Pupils will be in Year 10 for the beginning pre-testing and Year 11 for the intervention and post- testing.

Sample size – Approximately 120 pupils depending on which classes provide the best case-match.

Consent is not necessary as this trial would not be considered outside the natural remit of a teacher trying new teaching techniques. Pupils receiving the control will not be deprived of any teaching and will instead receive current best practice. Parents will be made aware of their child's involvement.

Parents will be able to opt out of having their children's data used in the analysis, and having anonymised data shared with third parties (the IEE). .

Assignment to condition

How will you allocate participants to the intervention and control groups?

Randomisation – achieved through a **clustered matched-pair design** by comparing data from

Year 10 teaching groups across the two schools. The primary measure to achieve this will be Key Stage 2 APS score- a measure used to determine streaming and relevant targets to the pupils. The measure will be used to select potentially similar groups for pre-testing. Four parallel sets will be case matched.

Once a best fit match has been carried out, then a pre-test will be administered to ensure there are no discrepancies in the match. The final match of Year 10 sets will be determined by Key Stage 2 APS and closest match of pre-test scores.

Once matched, the allocation will be random for the classes through a coin toss with heads denoting one set and tails the other. Whichever faces up after the toss will be the intervention group for Ashington, and therefore the other Ashington Year 10 set will be the control. This will work the same way for Bedlington.

Innovation

Describe your innovation in detail

This will be a single-blind trial where teachers are aware of whether their group is a control or intervention group. Pupils, however, will not be aware of whether they are in a control or intervention group. There will be a high level of mundane realism incorporated into the trial as the existing scheme of work is not being altered, and although LbQ is not an intervention which pupils have experienced before, they are used to receiving a short quiz at the start of each lesson as part of the SMART Connect retrieval practice which forms part of our lesson structure. Therefore, the likelihood of demand characteristics being encountered are low.

Control groups will be taught the standard scheme of work for science and teachers will use traditional SMART Connect retrieval activities consisting of six questions. They will not be given instant feedback on their performance. The scheme of work follows an accelerated learning model and provides a lesson outline and resources for teachers to deliver a short Connect (starter) activity, Activation, Demonstration and Consolidation activity each lesson. To ensure consistency with the intervention and no undue bias with exposure to multiple choice questions, the SMART Connect will also contain multiple choice questions in all lessons.

Intervention groups will be taught a modified scheme of work for science. Instead of the Connect activity each lesson, pupils will be given a short retrieval quiz via LbQ which consists of six multiple choice questions. They will be given instant feedback on their performance.

Outcome measures

What outcome measures will you use? When and how will they be administered and scored?

Science knowledge

A test comprising of 50 multiple choice questions will be created which will examine knowledge obtained in the units that have been covered during the trial. Each question will be worth one mark and therefore the test, like the pre-test, will be worth 50 marks. The test will be created by subject specialist staff who are not involved in the planning or implementation of the trial. The test will be developed on an Online Forms format and will self-mark to avoid teacher marking bias. The content of the pre-test will come from the subject matter of the upcoming units.

Before allocation, all groups will take the 50-question multiple-choice pre-test to determine baseline scores and aid in the most accurate case-matching. The same test will ultimately be used as a post-test. The length of time between tests (approx. six months) will negate any potential impact of re-testing using the same test. Teachers not involved in the trial will supervise the sitting of the pre-test to avoid teachers in the trial altering their normal teaching based on the contents of the pre-test.

The Resilience Scale

The Resilience Scale comprises 25 questions relating to confidence and the ability to be resilient. The Resilience Scale is a measure of a person's resilience in everyday life not just in learning. In each question the participant is asked to rate how they feel about the statement from 1–7, with 1 being strongly disagree and 7 being strongly agree. For example, the first statement is 'When I make plans, I follow through on them'. Learners would tick 1 if they strongly disagreed and 7 if they strongly agreed with 2–6 representing a scale in-between. The higher the numerical value obtained from The Resilience Scale the more resilient the participant reports themselves to be.

Pupils will complete the resilience scale the lesson before the pre- and post-tests to reduce the resilience being effected by knowledge of the tests. The trial is interested in the effect of LBQ on the resilience of learners due to the hypothesis that personalised private feedback which may aid in more risk-taking by the learner. In a study by Maes 2009, The Resilience Scale was found to be a 'valid and reliable tool' in testing resilience in a person.

Process analysis

What data will you gather for the process analysis?

Unobtrusive, brief observations will be undertaken once the trial is well underway to check on the undertaking of the trial to ensure that the correct methods are being followed and the control groups are not being exposed to any component of the intervention.

A brief questionnaire will be distributed after the trial to teachers. The questionnaire will touch on areas like how easy LbQ is to use, perceived enjoyment from pupils, any observable difference in resilience of pupils and effects on workload. The questionnaire will be created by teachers not involved In the test to prevent perceived bias.

Analyses

How will you analyse your outcome and process data?

Quantitative data obtained will be the pre- and post-test scores from the control and intervention groups and the scores from The Resilience Scale pre- and post-trial.

Both pre- and post-test will be a score out of 50 whereas the Resilience Score has a total score of 175, with 175 being the most resilient score.

It is expected that the trial will produce ordinal data with a normal distribution. As it is a between-participant pre- and post- test design the most fitting statistical test to consider the pre-test would be an ANCOVA. The resilience scale will produce ordinal data.

The trial would work off a p value of 0.05 to show significance. Effect size will be calculated as well as the median, mode and standard deviation of the pre-, post-tests and Resilience Scale scores.

Normal distribution will be ascertained using a Kolmogrov-Smirnov test. If the data is not normally distributed, a Mann-Whitney U test will be adequate for the data that is expected.

As sub- groups within the study, the scores of pupil premium pupils and that of boys and girls will be isolated for further discussion of the effect on those groups of learners.

Conclusion

Potential limitations

What are the potential limitations of your design?

The main limitation of this trial will be the sample size. A sample of 120 is relatively small and

therefore will be likely to inflate the effect size and reduce the chances of a significant result   .

Considering the small sample size, there is also a possibility of drop outs from the trial with set changes, natural fluctuation in pupil numbers and the relatively small chance of pupil exclusions.

Although measures will be taken to ensure similarity between the control and intervention group (KS2 APS and Pre-test scores) there are still possibilities they are not an exact match.

Since this trial will be set over two school this brings in other unique limitations. Even though the same scheme of work and content will be delivered, the trial will have little control over any variance in quality of teaching between schools and between classes. This could have a potential impact on the post-test scores. The small sample of pupils from two schools with similar contexts will reduce generalisability.

There is a possibility of bias occurring as the pre- and post-test will be the same test. However, there will be approximately six months between pre and post and therefore should negate some of this effect.

Both the pre/post-test and the intervention are technology based, which could potentially give the intervention group an unfair advantage. However, we are assuming a level of technical literacy which is generally found in this generation of children.

Implications

What would you do if your results were positive? What about if results are negative or neutral?

If the results showed that LBQ had a positive effect on recall then the next step would be to replicate the trial with a larger cohort to see if the results support the original. If the results do support, then this could possibly lead to a change in the memory retrieval practice within lessons to consider the new findings. The results would be taken to relevant decision-makers to make an informed decision about changes in practice in a whole school/trust level.

If the results appear relatively neutral, this could be down to limitations in the design. This would be cause to revise the trial and repeat to see if there is a difference in outcomes. If no change, then it is possible this is the actual result of the trial and the intervention would have to be changed for future research.

If the result appears negative then the effect is detrimental and the trial should not be repeated.

References

Karpicke J D and Blunt J R (2011), Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331(6018)*, 772–775.

McDaniel M A, Wildman K M and Anderson J L (2012), Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory & Cognition, 1*, 18–26.

Roediger HL and Karpicke J D (2006), Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.

Trumbo M C, Leiting K A, McDaniel, M A and Hodge G K (2016). Effects of reinforcement on test-enhanced learning in a large, diverse introductory college psychology course. *Journal of Experimental Psychology: Applied, 22*, 148–160