EEF Blog: Does research on ‘retrieval practice’ translate into classroom practice?

*EEF senior associate Prof. Rob Coe explains why, though there is lots of evidence to support the use of retrieval practice, there is still a question mark about how effectively it can be incorporated by teachers into lessons...*

For the first ever EEF Teacher Choices trial, we are working with 70 science teachers to compare the impact of starting a lesson with a retrieval quiz, or starting with a discussion to engage interest. Retrieval practice is strongly supported by over 100 years of research and is one of only two learning techniques rated by Dunlosky et al (2013) as having ‘high utility’ for classroom practice.

It is also growing massively in popularity in England. Indeed, some teachers we spoke to when planning the trial were concerned that we wouldn’t be able to find enough teachers who were not already using retrieval practice in the classroom. So, is it even worth evaluating: surely we already know that retrieval practice works?

Well, yes and no.

**Why am I not convinced that promoting retrieval practice will lead to better learning?**

There is a colossal amount of research to support the use of retrieval practice. It is true that, as with most research in psychology, this evidence primarily comes from laboratory studies with North American psychology undergraduates who get course credits for taking part.

But, although the majority of studies come from laboratory settings (223 vs 30 in classrooms according to the meta-analysis from Adesope et al, 2017), the effect sizes are similar in both (0.62 for lab studies vs 0.67 for classroom). The small number of studies conducted in primary schools (10 effects, mean 0.64) and secondary schools (19 effects, mean 0.83) are also comparable in their results to those in post-secondary settings (228 effects, mean 0.60).

These findings suggest that the current enthusiasm for retrieval is well-justified.

So why am I not convinced that promoting retrieval practice will lead to better learning?

First, because there are some outstanding questions about the types of learning best supported by retrieval. Many studies of retrieval focus on ‘relatively simple verbal materials, including word lists and paired associates’ (Dunlosky et al, 2013, p.32) and some cognitive scientists have questioned whether retrieval improves performance in complex tasks. Van Gog and Sweller (2015) argue that ‘... the testing effect decreases as the complexity of learning materials increases ... the effect may even disappear when the complexity of learning material is very high’ (p.247), while Rohrer et al (2019) note ‘benefits of retrieval practice have yet to be demonstrated for mathematics tasks other than fact learning.’

In Adesope’s meta-analysis (2017), the authors found that the 11 effects that required transfer are similar in size to those from retention (mean effect size 0.63 for retention, 0.53 for transfer). However, Agarwal’s recent paper (2019) provides some extra support for the thesis that we get better at what we practise, suggesting that the focus of retrieval questions matter.

But my biggest doubt is related to what Steve Higgins has called the Bananarama Principle: ‘it ain’t what you do it’s the way that you do it,’ (Higgins, 2018).

I think it is true that ‘to be able to retrieve, use, and apply knowledge in the long term, it is highly effective to practice retrieving, using, and applying knowledge during learning’ (Karpicke & Aue, 2015, p.318).
However, there is a big difference between demonstrating this in well-controlled, small-scale research studies in which experts in the ‘testing effect’ design retrieval activities and outcome tests and guide their use, and simply giving advice to typical teachers to incorporate retrieval quizzes into their lessons.

Why might the latter fail to work as the research says it should? Here are a few possible reasons:

1. Teachers might generate retrieval questions that focus solely on factual recall (these questions are easier to generate) rather than requiring any higher-order thinking.
2. Questions might be too easy and boost confidence without providing real challenge, which is likely to be a key ingredient for generating the kind of learning hoped for.
3. Teachers might allocate too much time to the quizzes, effectively losing the time they need to cover new material.

This list could certainly go on. The point is that avoiding these pitfalls (any one of which could prevent the ‘secure’ research finding that retrieval practice works from being demonstrated in real contexts) requires a mixture of skill (eg, being able to judge whether students have originally learnt the material, being able to create good questions), understanding (eg, that effects are biggest when recall is hard) and commitment (eg, making time to plan the quizzes and keep them going, reducing ‘teaching’ time to fit them in).

It seems likely that many teachers will not necessarily have all three.

If we can get a boost in student learning by giving teachers some simple guidance and encouraging them to follow it then our strategy is obvious: find out what works and share it widely.

If our advice is just to incorporate quizzing without support to build these capabilities, then it may well not work – despite all the research evidence that apparently supports retrieval practice. On the other hand, it may be that incorporating retrieval practice into lessons is actually relatively straightforward and that the prerequisites for making it work are either more common or less important than pessimists like me have assumed.

Which of these proves to be closer to the truth could make a lot of difference to schools and to those promoting teachers’ effective use of research evidence.

If we don’t get such a boost, things are a bit more complicated. Would clearer guidance have worked? Or perhaps effective quizzing requires more intensive training?

The EEF’s programme of work on Teacher Choices is designed to answer questions like these. Our first few trials are as much about investigating how teachers make choices and are able act on evidence as actually answering the impact question – no one has ever done these kinds of studies before and we have already learnt a lot about how complex they are!

Crucially, the independent evaluators from NFER have designed the trial so that we will learn about the kinds of barriers listed above: if it doesn’t work, we should get some good insights into which of the three reasons (or any others) might be the explanation.

These are the details that will determine whether the memories we retrieve from this period of English education are positive or negative.


