

RETRIEVAL PRACTICE AND THE MATHS ACHIEVEMENT GAP

Can a two-minute recall test close the gap between your weakest and strongest starters?

New evidence from first-year university mathematics courses, translated for secondary maths practice.

Which of your learners started this topic behind, and are they still behind by the end of the unit?

In one line

In two 13-week university maths courses, adding a short, unaided recall test at the end of every lesson was followed by the usual link between learners' starting knowledge and their results months later disappearing, while matched classes without the test kept that link firmly in place.



Starting knowledge stopped predicting results, but only with the recall test

Two first-year university maths courses for pre-service maths teachers, Number Theory (42 learners) and Algebra (50 learners), each ran for a 13-week term. All of them sat a diagnostic pre-test of school-level maths at the start. In matched practice groups covering identical lectures and content, one group solved two problems alone, unaided, in a short 5-10 minute test at the end of every lesson; the other group worked through the same two problems as a class, with the teacher's help. Months after each course ended (3 months for Number Theory, 5 months for Algebra), everyone sat a delayed post-test.

Without the recall test, starting scores kept predicting results ·

- ✓ Pre-test scores explained about 4 in 10 points of difference in the delayed post-test, in both courses ($R^2=0.40$)

With the recall test, that link nearly vanished · Pre-test scores

- ✓ explained under 1% of the difference in post-test results, in both courses ($R^2<0.01$)

Recall-test classes scored higher on the post-test itself ·

- ✓ Statistically significant in both courses (Number Theory $p=0.0001$; Algebra $p=0.012$)

The gap-closing pattern itself was statistically significant · Number

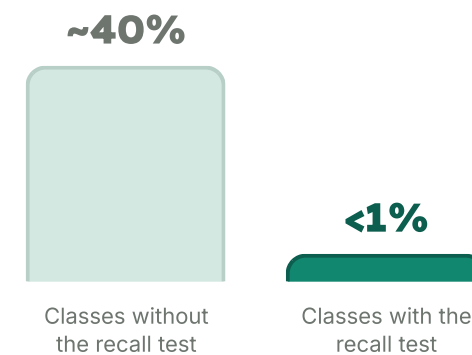
- ✓ Theory: $p=0.0012$; Algebra: marginally significant, $p=0.054$

It took months, not weeks, to show up · At the first mid-course test

- ✓ (six weeks in), the recall-test classes already scored higher, but the starting-knowledge link had not yet disappeared

How much of the gap in results traced back to where learners started

Share of the difference in delayed post-test scores explained by learners' pre-test scores (R^2), averaged across both courses.



Both courses (Number Theory and Algebra) showed close to this same pattern separately: $R^2=0.40$ without the recall test, $R^2<0.01$ with it, in each course.

A quasi-experiment with matched comparison groups, not a randomised trial: see page 4.

What the recall-practice groups actually did, lesson by lesson

THE MOVE

WHAT IT SOUNDS LIKE AT SCHOOL (OUR EXAMPLE)

WHAT THE STUDY DID

A two-question recall test, not a review

"Before you go, solve these two problems from today's lesson, on your own, no notes."

The recall-practice groups did exactly this at the end of every lesson, for 13 weeks straight: two open-ended problems on that day's content, similar to but not identical to what they had just practised.

SL GUIDANCE Do it well, for every learner: keep both problems tightly tied to that lesson's own content, so every learner has a fair shot.

Solved alone, not talked through

"Work it out yourself first; we will only go over it if people ask."

The comparison groups solved the identical two problems together as a class, with the teacher's help. That was the only planned difference between the two conditions.

SL GUIDANCE Do it well, for every learner: resist the urge to discuss the answers immediately. This study only discussed solutions if learners explicitly asked.

Marked lightly, fed back quickly

"Full marks, half marks for a small slip, or none. That is it."

The study used exactly this three-point scheme (1 point correct, 0.5 for a minor error, 0 otherwise) to keep marking fast enough for a teacher to sustain for a full term.

SL GUIDANCE Do it well, for every learner: a scheme this simple is quick enough to mark for every learner, every lesson, without it becoming a burden.

Judged over months, not weeks

"Let us see how this topic looks again after the holidays."

The gap between weaker- and stronger-starting learners had not closed at the six-week mark in this study. It only became clear 3 to 5 months later, on a delayed post-test.

SL GUIDANCE Do it well, for every learner: do not judge whether the gap has closed from a single end-of-unit test. Check again later in the term.

The honest fence: where this evidence stops

A quasi-experiment, not a randomised trial

Learners were split into practice groups by course-registration timing, not deliberately randomised to condition. The authors themselves say differing teaching styles "could not eliminate" this in their design.

University learners, not secondary-age learners

Participants were 18-23-year-old first-year university learners training to become maths teachers, studying Number Theory and Algebra, not secondary-age learners studying a typical secondary curriculum.

The key comparison used small numbers

In Algebra, only 14 learners in the recall-test group and 20 in the comparison group had complete data for the tightest test of the effect (the delayed post-test against their starting score).

Several different teachers were involved

Six teachers led Number Theory practice sessions and four led Algebra sessions. The teachers met weekly to align content, but the authors say plainly this "could not eliminate" differing teaching styles as a factor.

The authors do not credit recall alone

They state the benefit "isn't solely attributed to the act of retrieval itself". Extra feedback, test-taking practice, or engagement from any structured end-of-lesson task could also play a part, and was not measured separately.

Only one style of recall test was tried

Two open-ended problems, solved alone, is one specific format. Multiple-choice, flashcards or other recall formats were not compared in this study.

● Developing

A real matched-comparison design with statistically significant results converging across two separate courses, but not yet a randomised trial, and not yet tested with secondary-age learners. A randomised trial repeating this design with secondary-age learners, or independent replication by another research team, would raise this grade.

Worth testing: does a two-minute recall check close your own attainment gap?

The pattern worth borrowing is not Number Theory or Algebra, it is the structure underneath: a short, low-stakes, unaided recall task at the very end of the lesson, on that lesson's own content, kept up every lesson across a full term. That structure is just as easy to run in a secondary classroom as in a university practice session, and it cost the researchers no extra teaching time, only a few minutes of marking.

1 End the lesson with two questions, not a recap

Two problems from today's content, solved alone, no notes, in the last five to ten minutes. Keep it short enough that it never eats into new teaching.

2 Mark it fast, not hard

Use a simple right, part-right, or wrong scheme, like the study's 1, 0.5, 0, so feedback stays quick enough to sustain every lesson.

3 Track your weakest starters over a term, not a fortnight

In this study, the gap-closing advantage was not visible after six weeks, only months later. Follow your lowest-prior-attainment learners across a full unit before judging whether the gap has closed.

Developing signal, so run it as an experiment: pick a topic you teach across a term, add a two-question, unaided recall check at the end of every lesson, and compare your lowest-prior-attainment learners against the rest of the class at the end of the unit, not after the first fortnight.

Why not try? Planning a recall check into every lesson is easier with a framework to build it from. Open a free account to plan your next unit at app.structural-learning.com.