Description of timely practice

timely practice is a tool and collections of targeted teaching and learning resources for teachers of low attaining maths learners.

Within 33% of each maths lesson, learners complete their personalised PDF assignment and review the assessed assignment from their previous lesson, whilst their teacher gives personalised feedback when required.

The teacher uses the app to assess the assignments, and the app uses the data it collects to:

- show the teacher the firmness of learners' skills - so lesson planning and activities more frequently result in learning;
- schedule questions within assignments - so retrieval practice and feedback more efficiently embed learning.

Research Context

Learners with smaller working memories are more likely to be living with poverty (Farah et al 2006) and trauma (El-Hage et al 2006) and slow to learn in maths because they less reliably build chunks after the lesson (Gathercole 2008). Fortunately they benefit more from retrieval practice than their peers (Agarwal et al 2016). The mechanism appears to be that deliberate practice of similar but different problems helps build chunks (Gobet 2005)/mental schema in long term memory which reduces the working memory capacity required to solve problems (Sweller et al 1998).

timely practice was designed to significantly improve the proportion of teaching that becomes embedded learning in low attaining maths learners.

First teaching must become learning.

timely practice shows teachers the "firmness" for small bites of learning (layers) so teachers can more easily apply these research-backed raising attainment techniques:

- only teach skills for which the learner has mastered the pre-requisites (Kulik et al 1990, Wiliam 2009),
- teach the correct grain size (Sweller et al 1998),
- use a more spiralled curriculum (Gobet 2005).

Secondly learning must be embedded.

Learning is embedded most effectively using retrieval practice (Dunlosky et al 2013). The interval between one practice and the next should be:

- long enough to require effort to recall, but not so long that forgetting occurs (Bjork & Bjork 2011),
- increased after success (Kang et al 2014),
- reduced after failure and feedback given, ideally after a night's sleep (Pashler et al 2007).

timely practice automatically applies these recommendations so reteaching and overlearning (Rohrer & Taylor 2006) can be reduced, allowing groups of learners who are currently less successful at school than their peers, to be more successful than they currently are.

This research will compare, using standard tests, the attainment and motivation of classes using timely practice with those using traditional teaching.

Research Questions

Q1: What is the effect on the attainment of (currently) low attaining maths learners, when teachers and learners use timely practice in almost every maths lesson for at least one academic year, compared to a traditional teaching approach without using timely practice?

Q2: What is the effect on the motivation of (currently) low attaining maths learners, when teachers and learners use timely practice in almost every maths lesson for at least one academic year, compared to a traditional teaching approach without using timely practice?

Research Methods

Sampling plan

We would want to recruit as many schools as possible, but will limit our workload to 3 new schools per term. We will define low attaining learners as learners who are not expected to achieve a grade 4 or above in GCSE maths.

Learners will be assigned to the treatment or control groups by the maths class that they are in.

The controlled variables are:

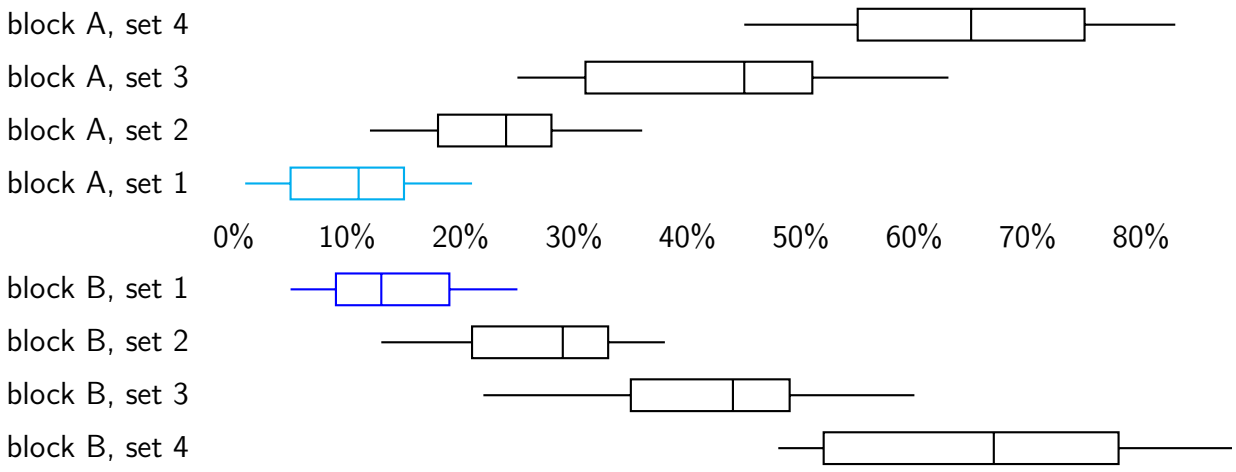
- classes are being taught towards the same exam,
- learners are in the same school year.

The manipulated variables are:

- the tightness of the spiral of the scheme of learning,
- the spacing of the practice questions,
- the availability of assessment data for the teacher to use to plan teaching.

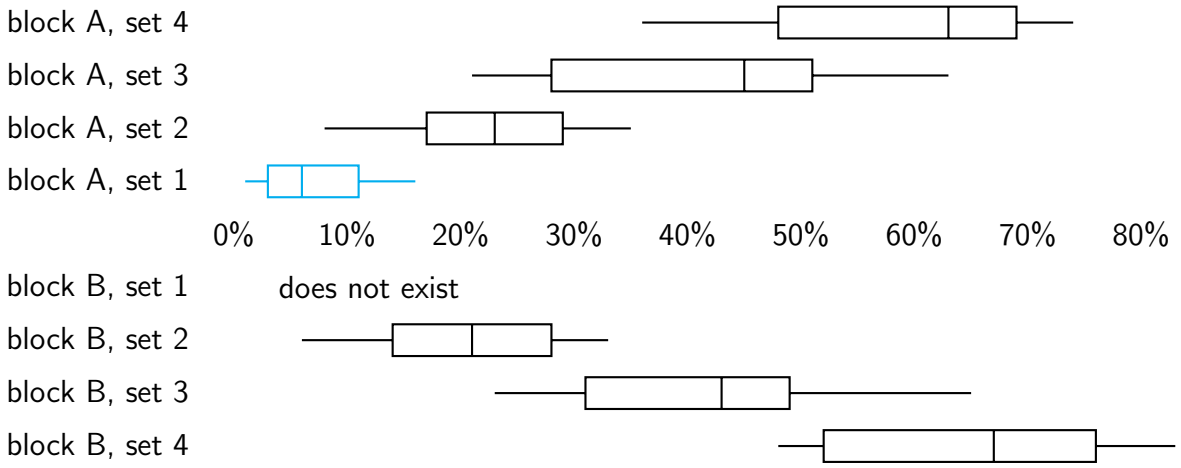
We wish to see if the treatment class, can significantly outperform expectations, and to do this we need to compare the treatment class not only with a control class that is similar in attainment, but also all the classes in the "set above".

Usually schools timetable in blocks and learners rarely move between time table blocks. Here is a school with 4 maths sets and 2 timetable blocks, where one of the two lowest attaining sets is going to be the treatment class.



The lowest attaining set, set 1, is made up of one class from block A and one class from block B either of these classes could be the treatment class. The other class from set 1 would then become **one** control class. There would be **two** additional control classes, both the classes from the "set above", that is set 2.

Some schools create time table blocks with all the lowest attaining maths learners in one timetable block. In this example, there is only **one** class in set 1, in block A. If the lowest attaining class is to be the treatment class, then the **two** control classes would be both the classes from the "set above", that is set 2.



The control and treatment classes would ideally:

(i) have learners randomly assigned to treatment or control class,

(ii) have the same maths teacher,

(iii) have their maths lessons at equally advantageous times of the day,

(iv) be equal in attainment and motivation (at the start of the trial),

(v) have equal numbers of learners.

However, we are working in the real world. (i), (ii), (iii) and (iv) are very unlikely because of blocking within school's timetables. (v) is very unlikely as the classes in "set above" above the treatment class will almost certainly have more learners than the lowest attaining sets.

For ethical reasons (see the next section) we have chosen:

- not to use the same tests across all schools,
- not to use individual test scores,

so we can readily compare classes progress within the same school but not between schools.

A minority of learners may be moved from the treatment to a control class (or v.v.). These learners' data will be excluded from the data analysis.

Data collection plan

For the attainment research question:

Since the control classes will almost certainly be following a depth first scheme of learning (being taught easy, middle and hard from a topic within a fortnight and not returning to the topic again until the next academic year), and the timely practice class should follow a breadth first scheme of learning (e.g. easy from each topic in term 1, middle from each topic in term 2, hard from each topic in term 3) the testing needs to be done carefully.

The trial shouldn't distract from the control classes' use of lesson time by excessive testing, see the ethical considerations section, so we will use a summative test created by the exam board for which the school is preparing their students. We expect this to be the GCSE foundation exam - it covers a wide mix of topics and difficulty, but will not be too time consuming to set or assess. Of course by using these tests we are limiting ourselves to a narrow definition of attainment.

The timing of the tests will be agreed in negotiation with each school, the minimum would be 1 complete comparative test set per year. Each complete GCSE exam is made up of 3 test papers. They cover year 1 to 11 of the national curriculum, so assessing part way through a year/two year GCSE course will make little difference to the coverage of the content tested.

We will compare the 5 figure values for each class before (in the middle of a two year course) and at the end of the trial. If we see that a lower attaining set's upper quartile moves from being below to above the lower quartile of the next more highly attaining set, then this is likely to be an indicator of closing the learning gap, rather than the expected "odd learner being ready to move up a set". We can then ask schools for similar data from previous cohorts (all the classes will now be control classes) to see how frequently this level of progress has previously occurred.

For the motivation research question:

We need to select a questionnaire which has been validated by researchers but is simple enough for low attaining learners to answer independently.

The questionnaire should be completed shortly before the trial begins and at the end of trial.

None of the following questionnaires are suitable without some modification:

Criterion	IMMS	RIMMS	AMS	EVC	PSMMS	BEAMMS
Take less than 30 minutes of lesson time	no	perhaps	no	yes	no	yes
Low reading age	no	no	no	yes	no	yes
Show small changes in motivation	yes	yes	perhaps	yes	no	no
Suitable for U.K. learners	no	no	no	no	no	almost

We propose to use the BEAMMS but:

- allow the 2 point scale to be expandable to a 6 point scale (see appendix 3) by allowing the learners to show if they slightly agree, agree or strongly agree with their chosen sentence. This will allow us to use the analysis as directed, but also see smaller changes in motivation.
- change some wording to make it suitable for U.K. learners. We will change the word **math** to **maths** and change the wording of the item 7 pair of sentences.

old 7.	Maths goes by pretty fast. ☐	Maths takes too long. ☐

Key:			
	☐	☐	strongly agree ■
	☐	agree ■	■
	slightly agree ■	■	■

new 7.	☐ Time seems to go quite quickly ☐ in maths lessons. ☐	☐ Time seems to go quite slowly ☐ in maths lessons. ☐

The standard BEAMMS analysis will result in a score out of 10, and the fine accuracy BEAMMS will result in a score out of 50.

We will compare the before and after 5 figure values of both scoring systems for all maths classes in the school. This will help us to see if there is a trend in the motivation scores due to age. We expect to see that learners from lower sets are less motivated than those from higher sets (Black and Wiliam 1998).

What timely practice attempts to do is increase engagement, learners doing more maths in maths lessons, so we will be particularly interested in items 4, 5 and 8 which are the items Orosco (BEAMMS 2016) says are indicators of engagement.

Ethical considerations

For the attainment research question

If we are to measure attainment we must test.

Our tests should measure embedded learning[1], yet take away no more motivation[2] nor time from learning[3], than the schools' existing testing regimen does. The tests must be fair to classes following different curricula[4], not be time consuming for the teachers to create and assess[5] and be considered a fair test by both ourselves and the school[6].

So the tests should be summative tests which the school intends to set anyway. Hence they will use little to no extra lesson time of the classes, will induce little to no extra demotivation and we will therefore avoid using the control classes to merely further our research.

For the motivation research question

We expect a 6 point scale as apposed a 2 point scale will enable us to see slight changes in learners' motivation. It may may also be less demotivating[7] and may lead to more honest answers[8].

---

[1] as learning which is soon forgotten and cannot be recalled is of no use, so there should be a delay of at least 6 weeks between when skills are learned and when the vast majority of skills are tested.

[2] high stakes testing increases demotivation in low attaining learners (Jacob 2001), but this is also true for less high stakes summative testing (Black and Wiliam 1998). Hence we should not increase the amount of summative testing occurring.

[3] to avoid loss of opportunity for learning because of the time required to take tests.

[4] The control classes learn the easy, middle and hard from each topic at one time, whereas the treatment class learns the easy in term 1, the middle in term 2 and the hard in term 3. So if we test what one class learned in the penultimate 6 weeks, it will be unfair to the other. Further, some of the control classes are learning harder curriculum content, so the tests should include what any learner my have learned, we must, to some extent, do summative testing. The school can make it as painless as possible by creating a "cut down paper"- a summative exam paper, with questions which are known to have not been learned by any learners, removed.

[5] using an existing exam paper would be ideal

[6] if we use exam board summative tests, or school set summative tests, the tests will be

- fairer,
- a better measure of embedded learning,
- less frequent and so take little lesson time,
- seen to be fair by learners, teachers and schools.

We can also expect the school to make sure these tests remain unseen. Most U.K. schools already use practice or mock GCSE tests in year 9 to 11 and many schools use these for years 7 to 11, so for many schools these will be the obvious choice.

[7] e.g. being asked to give more nuanced answer between *I feel bad when I do my maths work* and *I feel good when I do my maths work* will make the learner more likely to reflect, rather than jump to the more negative immediate response "I feel bad". Learners in secondary school will be well aware of what the "right" sentence is for each motivation assessment item (i.e. the answer the teacher wants to hear, the answer they think other more successful learners will give). Feeling forced to negatively rate themselves or lie are both likely to be demotivating.

[8] The learner will be more likely to consider "sometimes I feel bad, sometimes I feel good. I feel bad slightly more often than I feel good so, I'll slightly agree with the first sentence". This internal debate, in itself, may mitigate against demotivation.

Both research questions

In order to keep the learners' data private, and since with small classes the chance of identifying a learner is increased, this research proposes to compare the 5 figure values (minimum, lower quartile, median, upper quartile and maximum) of each class, so individual learners cannot be identified through the research.

Prior research indicates learners' outcomes can be expected to be improved through taking part in this research. There is no expectation on schools to continue the trial if they feel that learners are making worse progress than normal.

Schedule and Plan of Work

We will recruit as many schools as possible, and this will be time consuming. Not all schools need begin the trial/experiment at the same time. Schools can begin the trial at a number of points during the academic year, the constraint being when the school normally sets (or can change when they set) summative tests to the year groups involved in the trial.

Before starting the trial:

- ensure the school is happy to have the annual scheme of learning of the treatment class(es) cut and pasted into a new format,
- decide which will be the treatment class(es), this will then decide which teachers need to be trained and which classes will be the control classes,
- decide when teacher training will occur,
- agree when the comparative tests will be done,
- agree a start date for the timely practice class(es) to move to the adjusted scheme of learning,
- agree when the "before" and "after" motivation questionnaires will be done.

Additionally, in half term 0, before any changes are made to teaching and learning:

- some schools will wish to inform learners' parents of the treatment class(es), what will happen differently in maths lessons and why,
- do the "before" comparative test(s) or gather together suitable existing test scores,
- do the "before" motivation questionnaire.

Ideally, the "after" comparative tests should be done 6 half terms (one school year) after the trial begins. However, allowing between 5 and 7 half terms (for the one year trial) or 11 to 13 half terms (for a 2 year trial), between the "before" and "after" tests will allow the trial to fit in with the school's existing practice of testing e.g. if the school staggers when year groups take tests.

The "before" and "after" questionnaires should be done 6 half terms apart (or 12 half terms for a 2 year trial).

Analysis and write up will take a half term once the "after" comparative test data is available and the questionnaires are complete.

References

Each reference is accompanied by a summary - to make it quicker for teachers (who may not have the time or inclination to read each reference) to read this document.

Code	Summary Formal reference
Agarwal et al 2016	The benefits from retrieval practice are greater for students with lower working memory capacity. Agarwal, P.K., Finley, J.R., Rose, N.S., & Roediger H.R., Benefits from retrieval practice are greater for students with lower working memory capacity Memory 25(6):1-8 August 2016
AMS 1992	A motivation scale. Vallerand, R.J., Pelletier, L.G., Blais, M.R., Briere, N.M., Senecal, C. B., Vallieres, E.F., Academic Motivation Scale, High School Version, Adapted from AMS - College version 1992-1993 Educational and Psychological Measurement, vols. 52 and 53
BEAMMS 2016	A motivation scale. Orosco, M.J. Measuring Elementary Student's Mathematics Motivation: A Validity Study January 2016, International Journal of Science and Mathematics Education 14(5):945-958
Bjork & Bjork 2011	We embed learning more deeply, when we can recall it less readily. However we must not wait too long, if we have forgotten entirely then we will not be able to recall. Bjork, E. L., & Bjork, R. A.(2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds), Psychology and the real world: Essays illustrating fundamental contributions to society (pages 56-64) New York: Worth Publishers.
Black & Wiliam 1998	"The evidence is that the effect of (summative testing) is to teach the weaker pupils that they lack ability, so that they are de-motivated" Black, P. & Wiliam, D. (1998) Assessment and classroom learning, Assessment in Education, 5 (1), pp. 7-74.
Dunlosky et al 2013	Practice testing and distributed practice (often called retrieval practice) are the only 2 of the 10 techniques considered which are given their "high-utility" rating. Dunlosky, D., Rawson, K.A., Marsh, E.J., Nathan, M.J., Willingham, D.T., Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology, Psychological Science in the Public Interest, January 2013

	Summary Formal reference
El-Hage et al 2006	Learners living with trauma are likely to have a smaller working memory capacity than their peers. El-Hage, W., Gaillard, P., Isingrini, M., Belzung, C., Trauma-related deficits in working memory, February 2006, Cognitive Neuropsychiatry
EVC 2015	A motivation scale. Kosovich, J.J., Hulleman, C., Barran, K.E., Getty, S. A Practical Measure of Student Motivation: Establishing Validity Evidence for the Expectancy-Value-Cost Scale in Middle School, June 2015 The Journal of Early Adolescence 35(5):790-816
Farah, et al. 2006	Learners living with poverty are likely to have a smaller working memory capacity than their peers. Farah M.J., Shera, D.M., Savage, J.H., Betancourt, L., Giannetta, J.M., Brodsky, N.L., Malmud, E.K., Hurt, H. (2006) Childhood poverty: Specific associations with neurocognitive development. Brain Res 1110 : 166 - 174
Gathercole 2008	"The majority of children with poor working memory are slow to learn in the areas of reading, maths and science, across both primary and secondary school years". Learners with smaller working memories experience a double whammy of disadvantage - they are more reliant on chunks in long term memory to learn and yet are less likely to build these chunks in lessons. Learning difficulties arise because they "are unable to meet the memory demands of many structured learning activities" and "as a consequence, their working memory becomes overloaded ... information that is needed to guide the ongoing learning activity ... is lost". Gathercole, S. Working memory in the classroom, 2008, Presidents' Award Lecture at the Annual Conference of The British Psychological Society
Gobet 2005	"Use an 'improving spiral', where you come back to the same concepts and ideas and add increasingly more complex new information". Deliberate practice of similar but different problems assists the learner in building chunks in long term memory, which enables learning to be retained. Chunks also provide perceptual cues or triggers, so that the expert (or in our case the low attaining learner) can quickly and easily decide which chunk or chunks are likely to be useful to help solve a given problem. Gobet, F. (2005). Chunking models of expertise: Implications for education. Applied Cognitive Psychology, 19, 183-204.

Code	Summary Formal reference
IMMS 2015	A motivation scale. Loorbach, N. , Peters, O. , Karreman, J. and Steehouder, M. (2015), Validation of the Instructional Materials Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology. Br J Educ Technol, 46: 204-218.,
Jacob 2001	Although graduation tests have no appreciable effect on the probability of dropping out for the average student, they increase the probability of dropping out among the lowest ability students. Jacob, B.A. Getting Tough? The Impact of High School Graduation Exams, June 2001 Educational Evaluation and Policy Analysis 23(2):99-121
Kang 2016	When learners do repeated retrieval practice attainment is raised. Kang, S.H.K. (2016) Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction. Policy Insights from the Behavioral and Brain Sciences, 3 (1), 12-19
Kang et al 2014	Expanding interval, as opposed to fixed-interval, retrieval practice is more efficient. Retrieval practice over the long term: Should spacing be expanding or equal-interval? Kang, S.H.K., Lindsey, R.V., Mozer, M.C., and Pashler, H. Psychonomic Society, Inc. 2014
Kulik et al 1990	When teachers hold back from teaching new learning until the learners have mastered all the pre-requisite skills, attainment is raised. Effectiveness of mastery learning programs: A meta-analysis Kulik, C.C., Kulik, J.A., & Bangert-Drowns, R.L.
Pashler et al 2007	Giving feedback when the learner can't retrieve (recall the learning) makes retrieval practice more effective. Feedback after a night's sleep is more effective than immediate feedback. Pashler, Rohrer, Cepeda & Carpenter (2007). Enhancing learning and retarding forgetting: Choices and consequences. Pyschonomic Bulletin & Review 2007, 14 (2) 187-193
PSMMS 2015	A motivation scale. Ersoy, E. and Oksuz, C., Primary School Mathematics Motivation Scale, European Scientific Journal June 2015 edition vol.11, No.16 ISSN: 1857 - 7881

Code	Summary Formal reference
RIMMS 2015	A reduced question version of the IMMS motivation scale. Loorbach, N. , Peters, O. , Karreman, J. and Steehouder, M. (2015), Validation of the Instructional Materials Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology. Br J Educ Technol, 46: 204-218.,
Rohrer & Taylor 2006	Overlearning, getting learners to practise more questions once the learners have "got the skill", adds little, if any durability to learning. Rohrer, D. & Taylor, K. (2006) The effects of overlearning and distributed practice on the retention of mathematics knowledge. Applied Cognitive Psychology, 20, 1209-1224.
Sweller et al 1998	When teachers teach the "right grain size" attainment is raised. "Schema construction has two functions: the storage and organization of information in long-term memory and a reduction of working memory load." People have very limited working memory capacities so attempting reasoning in working memory is inefficient and error prone. Whereas prior problem solving, which helps us create schema, in turn helps us to automate future problem solving - giving us more working capacity for the unfamiliar parts of a problem. Even a very complex schema can be used by working memory as a single element. Building and using an increasing number of ever more complex schemas, by "combining elements of lower level schemas in long term memory" allows skilled performance to develop. Sweller, J., van Merrienboer, J.J.G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10, 251-296.
van de Pohl et al 2010	Teachers should fade scaffolding after teaching. The duration over which the "fade" should occur is not quantified, but this writer thinks during the course of a single lesson is too fast for many low attaining learners. van de Pohl, J., Volman, M. & Beishuizen, J. Educ Psychol Rev (2010) Scaffolding in Teacher-Student Interaction: A Decade of Research 22: 271
Wiliam 2009	When teachers find out what learners know before and during teaching and alter their teaching in response to this information, attainment is raised. Assessment for Learning: why, what and how, Dylan Wiliam, An Inaugural Professorial Lecture, 2009, Institute of Education, University of London

Appendix 1: Training teachers

The focus here is allowing teachers to accept that their current teach-learn regimen may not be very effective for low attaining learners and to share in a little more detail what research tells us about embedding maths learning.

When we measure learning (i.e. do assessment) can have a profound effect on what we find e.g. if you think of what you taught last lesson to the kind of class who will be in the trial, what proportion of the learning did/will the learners be able to recall

- at the end of the lesson?
- after a week?
- after a month?
- after a year?

If we find the proportion of learning when we measure learning at the end of the lesson and after a year are very similar, then we can be happy that an each-topic-once-a-year scheme of learning and leaving learners to be in charge of their own revision will work well enough.

However if we find that when we measure learning at the end of the lesson and after a week and after a month, the proportion of learners who can recall and use the learning of the lesson significantly decreases, we may find that for leaners who can't or won't revise for themselves, that an each-topic-once-a-year scheme of learning is not appropriate.

Let's look at what Rohrer and Taylor found about how well learners could recall a lesson on the number of permutations (when some of the items are repeats) to college students in the US (who would be called university students in the UK).

In experiment (A) **when** learners were assessed was varied - they scored:

- at the end of the lesson: 95%
- after a week: 66%
- after a month: 25%
- after a year: no test data.

In experiment (B) the **number of practice question** learners completed in the lesson was varied. When learners were assessed after a week (or a month) - they scored:

- 3 practice questions in the lesson: 64%, (26%)
- 9 practice questions in the lesson: 66% (27%)

So for both experiment (A) and (B) what proportion of the learning for "learners who can't or won't revise" would you expect to survive for at least a year?

If, for "learners who can't or won't revise", we schedule their revision within their maths lessons, we may be able to increase the amount of learning that the learners can recall and use, provided that the revision takes less time than the lesson time wasted by overlearning and forgetting.

timely practice was designed to make, "making changes to what happens after the lesson", both easy for the teacher to do and far more effective than in Rohrer and Taylor experiments in embedding learning. In their experiments, by waiting at least a week between teaching and first practice, at least 30% of learning was lost. We can expect for low attaining learners the loss within a week might be greater, so by scheduling practice sooner than a week we might hope to avoid re-teaching as much as possible.

Of course, we as teachers, might criticise Rohrer and Taylor's experiment in that

- they didn't assess the learners in advance of the lesson to see if they had all the pre-requisite skills,
- they may have taught too much at one time,
- they may have chosen (by accident or design) to teach a skill which learners quickly forget,
- they didn't give the learners homework, nor tell them they should revise for a test.

When you use timely practice what you will do will be different than in Rohrer and Taylor's experiment (A) in that

- we will encourage you to use assessment for learning data: to ensure you teach in the sweet spot between too easy and too hard,
- we will encourage you to only teach a small bite of learning per topic per term: so that you are unlikely to teach too much,
- you will teach topics from your school's scheme of learning,
- rather than just assessing the learner twice, (once directly after teaching, once after a week or a month), or three times (once directly after teaching, once after a week and once after a 2 or 5 weeks) as Rohrer and Taylor did, you will assess the learner more frequently to start with. You will assess the learning at intervals which timely practice judges will extend the duration of learning of the lesson for longer each time, and so you can be more sure that e.g. what you teach at the start of year 10 (or indeed in year 8), will be able to be independently recalled and accurately applied within the GCSE exam towards the end of year 11.
- we will ask you to give feedback when the learners need it, (so that skills which are partially learned can be fixed) and the app will schedule skills that have required feedback soon (so that feedback assists the learner learn and is not wasted by the learner forgetting the feedback),

- but we will encourage you to sometimes judge some of the teaching of your lessons which can't be recalled "best learned later", when you find that the feedback for some bites of learning is taking too much of the teacher's and/or the learner's time.

Of course your objection to this might be "where will I find the time to do all this extra practice and feedback, I can't even cover the curriculum as it is?". However suppose your head of department or head teacher said

- "I want you to make sure that the vast majority of what you teach in one year can be recalled and used by the learners next year, so that the learners can nearly always learn new harder work next year on the topic. I'm happy if you teach less, in the hope that the learners retain more."
- "I want you to reduce the time you spend on practising similar questions in the lesson, to allow more time for retrieval practice in later lessons, just as Rohrer and Taylor did in experiment (B)"

Would you agree to try these changes to the

- scheduling of assessment, practice and feedback,
- fraction of the annual scheme of learning on a topic taught per term?

We will train the teachers by

- demonstrating (practically),
- sharing with teachers what academic research tells us about how teaching becomes long term learning, through conversation, presentations and videos,
- feedback on how the class is progressing, ideally every half term,

When possible we will train the teachers using timely practice - so they can experience how it works - and how it might save learners lesson time, which can be used for teaching more and harder.

We will be available for telephone/visits as suits the teachers.

Appendix 2: Schools that don't set by maths attainment

Although these schools are in the minority, they are often alternative provision schools - and often have learners who can be classified as underachieving - so including them in these experiments is desirable. In general, these schools will have only one class in each year, therefore it will be harder to find a control group. Comparing the timely practice class with a control class in the previous or subsequent year (providing none of the classes are in year 11) is one possibility. The other possibility is that the teacher would divide the class into timely practice group and a more traditional practice group.

The timely practice group will be made up of the lowest attaining learners (and perhaps also higher attaining learners) and the more traditional practice group will be made up of learners from perhaps the 40th to the 80th percentile. When the teacher allocates lesson time for the treatment (timely practice) group to do their timely practice assignments and get feedback, the control group will do warm up work, teacher selected extra practice questions or short tests. The practice questions can be from the topic of the lesson or from previous lessons.

The class would need at least one teaching assistant so that during the part of the lesson that the treatment group are doing their timely practice assignments:

- one teacher/teaching assistant will work with the treatment group: keeping them on task and giving help and feedback on the timely practice assignments,
- the other teacher/teaching assistant will work with the control half of the class: keeping learners on task, and giving any help or feedback that is required by the control group as they do their allocated work.

If the teacher and the teaching assistant have very different maths teaching competencies, then this could skew results unless e.g. the teacher and teaching assistant alternate which group they work with during the timely practice/extra practice part of the lesson. Ideally the teacher/teaching assistant would swap roles no more frequently than every half term, since giving feedback is best done by the same person from one lesson to the next.

Appendix 3: Beliefs, Engagement, and Attitude Maths Motivation Scale (BEAMMS)

Here are 10 pairs of sentences.

You **must** choose either the left or the right hand sentence.

Use this key to show how much you agree with each of your choices.

☐ ☐ slightly agree ■	☐ agree ■ ■	strongly agree ■ ■ ■

1.	I like doing maths.	I do not like doing maths.
2.	Doing maths makes me sad.	Doing maths makes me happy.
3.	I feel good when I do my maths work.	I feel bad when I do my maths work.
4.	I like spending my energy doing maths.	I do not like spending my energy doing maths.
5.	When I'm doing maths, I would rather be doing other things.	When I'm doing maths, it's what I want to be doing.
6.	Maths is easy for me.	Maths is hard for me.
7.	Time seems to go quite quickly in maths lessons.	Time seems to go quite slowly in maths lessons.
8.	I get bored when I do maths.	I get excited when I do maths.
9.	I do not care how well I do in maths, maths is not important.	I really want to do well in maths, maths is important.
10.	I choose to do maths because I like it.	If I had a choice I would not do maths.

The scoring depends on whether the learner chooses the more motivated or the less motivated sentence.

With the fine accuracy BEAMMS the scoring will be as follows:

more motivated sentence:	☐ ☐ 3 points ■	☐ 4 points ■ ■	5 points ■ ■ ■
less motivated sentence:	☐ ☐ 2 points ■	☐ 1 point ■ ■	0 points ■ ■ ■

Example answers to items 8, 9 and 10 are shown below

8.	☐ ☐ I get bored when I do maths. ■	☐ ☐ I get excited when I do maths. ☐
9.	☐ ☐ I do not care how well I do in maths, maths is not important. ☐	■ ■ I really want to do well in maths, maths is important. ■
10.	☐ ☐ I choose to do maths because I like it. ☐	☐ If I had a choice I would not do maths. ■ ■

For the fine accuracy BEAMMS these answers would score $\frac{8}{15}$ because

- for item 8: the learner slightly agrees with the less motivated sentence → 2 points,
- for item 9: the learner strongly agrees with the more motivated sentence → 5 points,
- for item 10: the learner agrees with the less motivated sentence → 1 point.

Whereas for the standard BEAMMS these answers would score $\frac{1}{3}$ because

- for item 8: the learner selected the less motivated sentence → 0 points,
- for item 9: the learner selected the more motivated sentence → 1 point,
- for item 10: the learner selected the less motivated sentence → 0 points.

BEAMMS was developed for primary maths learners. In the BEAMMS trials (BEAMMS 2016), the learners were in grade 2 or 3 - equivalent to UK years 3 or 4. The number of more motivated sentences a learner chose was used to split the learners into 3 groups the less, mid and more motivated groups, who endorsed (0 to 6), (7 - 8) and (9 -10) of the motivated sentences.

Below is a summary of the proportion of each sentence that the less and more motivated groups endorsed. The proportion who chose the less motivated sentence is shown (in brackets)[9].

First the 61 **less** motivated maths learners

1.	I like doing maths.	48%	I do not like doing maths.	(52%)
2.	Doing maths makes me sad.	(43%)	Doing maths makes me happy.	57%
3.	I feel good when I do my maths work.	66%	I feel bad when I do my maths work.	(34%)
4.	I like spending my energy doing maths.	6%	I do not like spending my energy	(94%)
5.	When I'm doing maths, I would rather be doing other things.	(98%)	Maths is the only thing I want to do.	2%
6.	Maths is easy for me.	57%	Maths is hard for me.	(43%)
7.	Maths goes by pretty fast.	33%	Maths takes too long.	(67%)
8.	I get bored when I do maths.	(92%)	I get excited when I do maths.	8%
9.	I do not care how well I do in maths, maths is not important.	(31%)	I really want to do well in maths, maths is important.	69%
10.	I choose to do maths because I like it.	26%	If I had a choice I would not do maths.	(74%)

Now the 70 **more** motivated maths learners.

1.	I like doing maths.	99%	I do not like doing maths.	(1%)
2.	Doing maths makes me sad.	(0%)	Doing maths makes me happy.	100%
3.	I feel good when I do my maths work.	99%	I feel bad when I do my maths work.	(1%)
4.	I like spending my energy doing maths.	91%	I do not like spending my energy	(9%)
5.	When I'm doing maths, I would rather be doing other things.	(20%)	Maths is the only thing I want to do.	80%
6.	Maths is easy for me.	94%	Maths is hard for me.	(6%)
7.	Maths goes by pretty fast.	99%	Maths takes too long.	(1%)
8.	I get bored when I do maths.	(1%)	I get excited when I do maths.	99%
9.	I do not care how well I do in maths, maths is not important.	(0%)	I really want to do well in maths, maths is important.	100%
10.	I choose to do maths because I like it.	99%	If I had a choice I would not do maths.	(1%)

Orosco says items 4, 5, and 8 are indicators of engagement and items 1, 2, 3, 6, 7, 9, and 10, are indicators of beliefs and attitudes, although he doesn't say which are which.

[9] It wasn't clear from reading the research paper, if the learners always selected a sentence from each pair, so this may include some learners who did not endorse either item.

What is clear, even at this young age, is that low motivation more strongly affects the engagement items than the attitude and beliefs items: fewer than $\frac{1}{6}$ of the less motivated learners endorsed each engagement item, whereas at least $\frac{1}{4}$ of the less motivated learners endorsed each belief and/or attitude item.

