



Education
Endowment
Foundation

Thinking Mathematically

Pilot report

June 2026

Jonah Bury, Sara Bashir Malik, Ellen Smith, Lydia
Marshall



Oxford
MeasurEd




The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries, and colleges to improve teaching and learning for 2–19-year-olds through better use of evidence.

We do this by:


- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

For more information about the EEF or this report please contact:

 Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

 info@eefoundation.org.uk

 www.educationendowmentfoundation.org.uk

Contents

About the evaluator	3
Executive summary	4
Introduction	6
Methods	16
Conclusion	66
References	74
Appendices	78

About the evaluator

This evaluation was conducted by a team from Oxford MeasurEd, including Dr Lydia Marshall (Principal Investigator and Project Director), Dr Jonah Bury (Senior Evaluator and Project Manager), Sara Bashir Malik (Junior Evaluator and Project Coordinator), Robert Wishart (Impact Evaluation Expert), Paulina Valenzuela (Junior Evaluator and Assessment Expert), and Ellen Smith (Junior Evaluator).

Oxford MeasurEd is a boutique global education consultancy dedicated to working with partners to collect and analyse data and improve learning for all children around the world. Our work focuses on designing and evaluating learning assessment systems, evaluating programmes, and providing strategic and capacity strengthening services to a range of education actors.

Contact details

Lydia Marshall (Principal Investigator)
Oxford MeasurEd
27 Mortimer Street
London W1T 3BL
Lydia.marshall@oxfordmeasured.co.uk

Executive summary

The project

The Thinking Mathematically (TM) programme aims to improve pupils' maths attainment and narrow the attainment gap between pupils from disadvantaged backgrounds and their peers by improving mathematical problem-solving skills. It does this by equipping teachers with the knowledge and skills to teach mathematical problem-solving effectively and consistently. The programme is a whole-cohort intervention aimed at all Year 7 children (11–12-year-olds). In-school delivery is overseen by one of the two lead teachers at each school (a 'delivery lead'); lead teachers are experienced maths practitioners. TM problem-solving sessions are delivered to all Year 7 maths classes by their regular maths teacher once per week for 20 weeks. In these 20-minute sessions, teachers model their own thinking using a 'think aloud' method and worked examples based on a pre-selected problem. Teachers then provide pupils with support using prompts and create debrief opportunities for pupils to discuss their methods and thought processes.

The TM programme was created by Simon Cox, the Director of Blackpool Research School (BRS). BRS is part of the Education Endowment Foundation's (EEF) Research School Network. A team from BRS delivers in-person training at regional locations to two lead teachers per school, who then 'cascade' the learning to Year 7 maths teachers. There are two 'touchpoints' for lead teachers to revisit key aspects of the teaching approaches throughout delivery. Targeted support is provided to a small sample of schools via in-person visits, while all schools have a short one-to-one online meeting with the delivery team to share experiences. This complements implementation data that the delivery team receive from lead teachers via delivery logs.

A total of 197 participants (55 lead teachers and 142 Year 7 teachers) from 28 schools in the North West and Yorkshire and Humber took part in the TM pilot between September 2024 and April 2025. The pilot evaluation used a mixed-methods design and included qualitative interviews with the delivery team and lead teachers; focus groups with Year 7 maths teachers; pre- and post-delivery surveys with teachers, and interviews with pairs of pupils. The evaluation team also carried out desk-based research and analysis of programmatic data (attendance and monitoring data) collected by the delivery team. Finally, the evaluation team observed in-person training sessions, online regional check-in meetings, and delivery of TM sessions. All qualitative research with schools, including the observations of classroom sessions, took place with ten purposively selected 'case study' schools.

Table 1: Summary of pilot findings

Area of research	Findings
Feasibility Is the approach feasible to implement?	The programme was partly delivered as intended. While training took place as planned and cascading took place in most schools, the programme was not consistently delivered as intended, with variation regarding session duration and coverage. Classroom sessions were delivered regularly, although teachers did not consistently adhere to the core components, with inconsistent usage of programme materials and strategies. Teachers praised the classroom resources for their clarity but highlighted that some questions were less suitable for pupils with special educational needs and disabilities (SEND) and English as an additional language (EAL). To address challenges to fidelity caused by inconsistent cascading, the delivery team has considered tightening up the cascading model, including communicating the requirements of cascading more clearly to senior leaders. To ensure mathematical problems are suitable for different pupil groups in the future, the delivery team are considering including differentiation, providing different difficulty levels for the same questions.

<p>Evidence of promise</p> <p>Is there evidence to support the theory of change (ToC)?</p>	<p>Findings indicate that the programme has some evidence of promise, but that some causal pathways are weaker than others. Although the response rate to the post-delivery surveys was low, most respondents reported improved problem-solving for pupils as a positive outcome of the programme. Pupils and teachers perceived problem-solving to be different to ‘business as usual’ teaching, and pupils expressed being motivated to engage with the programme. However, teachers reported that pupil engagement was not consistently high and was shaped by how accessible and relevant pupils found the problems. Teachers were motivated to engage with the programme, although their engagement was negatively affected where pupils were disengaged. Pre-post analysis shows positive shifts in teachers’ understanding of and confidence in teaching problem-solving. However, when asked specifically about benefits resulting from the programme, only a small majority of teachers reported clear personal gains. Those teachers who identified improved confidence in interviews noted that the routine of teaching problem-solving following the same structure throughout the programme helped develop confidence. Post-delivery survey data indicates promising levels of sustainability. Nearly three-quarters of teachers reported intending to sustain key practices, including encouraging pupils to discuss and compare approaches (the debrief) and modelling their thinking through ‘think aloud’. However, it is important to note that the self-selecting nature of the respondents means these survey results may not represent the views of all pilot participants.</p>
<p>Readiness for trial</p> <p>Is the approach ready to be evaluated in a trial?</p>	<p>The training was not consistently and effectively cascaded to Year 7 teachers by lead teachers; given the importance of cascading in ensuring implementation fidelity, the intervention is not yet fully ready to be evaluated for a trial and would instead benefit from a further pilot evaluation. The programme is partially manualised: comprehensive and high-quality materials exist but require additional guidance and clearer articulation of core components before delivering the programme to a larger number of schools, including a more explicit distinction between non-negotiable and flexible elements. To ensure fidelity at a larger scale, the delivery team would benefit from a more refined system for monitoring all core components.</p> <p>The evaluation team identified the Progress Test in Maths (PTM) as an appropriate outcome measure for an efficacy trial. The problem-solving subscale raw score can be used as a secondary, exploratory outcome, as it aligns with the programme’s focus on problem-solving. PTM is UK normed and offers practical whole-class administration with centralised scoring.</p>

Additional findings

As well as being willing to continue with key practices, there is some evidence of the programme leading to a more consistent departmental approach to teaching problem-solving. Among teachers who responded to both the pre-delivery and post-delivery survey, around two in three of those who initially reported having ‘not at all’ or ‘only slightly’ consistent approaches to teaching problem-solving at their school moved to reporting ‘consistent’ or ‘very consistent’ approaches after the programme. However, given the self-selecting nature of respondents, these findings should be interpreted with caution and may not reflect the views of all participating teachers.

Positive and negative unintended consequences materialised during programme delivery. There was evidence of teachers using the approach with other year groups, thereby potentially improving problem-solving outside of the Year 7 cohort. The programme was also perceived to have improved pupils’ communication skills, with two in three teachers responding to the post-delivery survey reporting improved communication as a benefit (see also EEF, 2022b; Mercer and Sams, 2006). In qualitative encounters with the delivery team, pupils, and teachers, the debrief element of the TM sessions was particularly highlighted as improving pupil oracy. However, participants also reported more negative unintended consequences. For instance, just under half of teachers (45%) reported reduced time available for other curriculum content as a negative by-product of programme participation. There was also a perception that the programme risked positioning problem-solving as a standalone aspect of the curriculum, potentially depriving pupils of problem-solving opportunities outside the weekly TM session.

Introduction

Background evidence

The Thinking Mathematically (TM) programme responds to evidence showing that pupils, and particularly those from disadvantaged backgrounds, underperform in mathematical problem-solving due to limited exposure, inconsistent teaching approaches, and gaps in metacognitive strategy use. While many pupils leave primary school without achieving the expected level in maths, a persistent attainment gap exists between socio-economically disadvantaged pupils¹ and their less disadvantaged peers. For instance, in 2023, 59% of pupils from disadvantaged backgrounds reached the expected level in maths at the end of Key Stage 2 (KS2) compared to 79% of pupils from non-disadvantaged backgrounds (Boylan *et al.*, 2024).

In KS2 and GCSE standardised national assessments, all pupils, but particularly those from disadvantaged backgrounds, tend to score less well on problem-solving questions than on fluency questions (Russell, 2000; Unity Research School, 2025).^{2,3} This may be partly because pupils in UK secondary schools have limited opportunities to practice problem-solving in class (Boylan *et al.*, 2024). Beyond the immediate application in maths, problem-solving skills are increasingly being valued by employers (Maths Horizons, 2025) and recognised as part of broader learning skills, along with communication skills (see Mercer and Sams, 2006), teamwork, and resilience. These skills are seen to be relevant to Britain's economic recovery and social cohesion in the context of regional disparities in skills and economic growth (The Council of Skills Advisors, nd). However, schools often do not have a pre-existing consistent approach to the teaching of mathematical problem-solving (Noyes *et al.*, 2023; Ofsted, 2023) or the time and expertise across their maths department to select suitable problems that will support pupils in developing problem-solving proficiency (Leavy and Hourigan, 2020; Walsh, 2016). Moreover, schools provide limited continuous professional development (CPD) opportunities focused on metacognition in the context of maths (Boylan *et al.*, 2024), which is central to teaching problem-solving. While some metacognitive elements appear in broader CPD programmes, such as the National Centre for Excellence in the Teaching of Mathematics (NCETM) mastery approach, sustained PD with metacognition as its central focus is uncommon (Boylan *et al.*, 2024). The TM programme, which is focused on the explicit teaching of problem-solving strategies, has not been available for use by schools except as part of this evaluation and early rollouts of the programme.

A secondary maths practice review by Boylan *et al.* (2024) found that teachers view problem-solving as a key part of maths, typically integrating it across most topics and teaching specific problem-solving strategies, rather than using extended tasks or whole lessons devoted to problem-solving (Boylan *et al.*, 2024). The current national curriculum for mathematics states that problem-solving is one of the three key areas of mathematics, along with mathematical fluency and reasoning, and that pupils should be explicitly taught to solve problems in KS3 (Department for Education, 2021). Moreover, the recently published Curriculum and Assessment Review underlined the importance of allowing for 'non-routine problem-solving that stretches all pupils', particularly across KS2 and KS3 (Curriculum and Assessment Review, 2025).

Recent polls with maths experts reveal that problem-solving is felt to be a particular area of weakness in the curriculum (Noyes *et al.*, 2023), with secondary maths leaders highlighting problem-solving as a key professional development priority (Boylan *et al.*, 2024). The literature suggests that this is due to (1) teachers making assumptions that structured argument and proof are only achievable by higher-attaining pupils, (2) the design of assessments which undervalue problem-solving through mark allocations that incentivise teachers to focus on memorising tips and tricks at the expense of building fundamental knowledge, and (3) the crowded curriculum (Maths Horizons, 2025).

There is some evidence to suggest that disadvantaged pupils are less likely to use metacognitive and self-regulatory strategies without being explicitly taught these strategies (EEF, 2018; Crenna-Jennings, 2018). Providing regular opportunities for pupils to explore various strategies for problem-solving can enhance their confidence, efficiency, and

¹ We define disadvantaged pupils as those pupils eligible for free school meals (FSM).

² This information is based on data available to schools from national assessment providers.

³ Fluency questions tend to assess pupils' accuracy with core procedures (for example, adding fractions). In contrast, problem-solving questions require pupils to choose and apply mathematical strategies in unfamiliar or complex situations. While fluency builds secure foundations, problem-solving develops reasoning, flexibility, and metacognitive skills.

flexibility in choosing appropriate approaches (Woodward *et al.*, 2012). It helps pupils persevere despite making mistakes, therefore increasing their mathematical independence and resilience (Ofsted, 2021).

The Thinking Mathematically (TM) programme sets out to address these challenges. For Year 7 pupils, the aims of the programme are to (1) better equip them to tackle unfamiliar mathematical problems, (2) narrow the disadvantage gap between pupils from disadvantaged backgrounds and their peers at Key Stage 3 and GCSE, and (3) improve maths attainment. For participating Year 7 maths teachers, the programme aims to equip them with the knowledge and skills to adapt teaching to explicitly address problem-solving challenges.

TM has several core components backed by evidence, including:

- **Use of a bank of problem-solving tasks**—The programme resources include a bank of problem-solving tasks that pupils do not have a well-rehearsed, ready-made method to solve, as recommended by the EEF Guidance Report on improving maths (EEF, 2022b; see also Lester, 2013).
- **'Think aloud' approach**—Encouraging teachers to 'verbalise their metacognitive thinking as they approach and work through a [mathematical problem]' (EEF, 2018) to support the development of pupils' problem-solving skills. Teachers do this by modelling how to plan their solution to a problem, monitoring their progress towards the solution, and evaluating the process of solving a problem while reflecting aloud. Teachers are encouraged to use worked examples to illustrate how they approached a mathematical problem step-by-step (Perry *et al.*, 2021).
- **Use of heuristics**—Providing pupils with a list of prompts (see Appendix 6) to help them monitor and reflect during the problem-solving process can enable them to develop a habit of mathematical talk (Woodward *et al.*, 2012). These prompts are an adaptation of the EEF's Thinking Aloud planning tool (EEF, n.d.).
- **A debrief**—Facilitating a debrief in which pupils are asked to explain each step used to solve a problem can enable them to communicate their reasoning and choice of strategy as well as compare and reflect on other strategies used by their peers (Hiebert and Grouws, 2007; Woodward *et al.*, 2012).

The TM pilot builds on two previous small-scale rollouts of the programme as part of the EEF's early-stage programme development work. The first rollout took place in Cumbria and North East England and involved ten secondary schools with approximately 250 Year 7 pupils, with one Year 7 class per participating school. The second rollout took place between April and July 2023 in the North, North East, and North West of England and involved a further ten schools.⁴ Unlike with the first rollout, all Year 7 pupils in participating schools took part (approximately 1800 pupils overall). In the second rollout, schools were required to deliver the programme to Year 7 pupils at least once a week over a 10-week period. Half of all participating schools came from Education Investment Areas (EIA)⁵ with low educational outcomes. While free school meal (FSM) eligibility was not a specific focus or requirement for taking part, participating schools from EIAs featured a higher-than-average proportion of pupils eligible for FSM. The delivery team collected the data and presented the evidence from the attendance records and survey responses from the second rollout in a report to the EEF, showing that the programme was acceptable and feasible in schools. Key findings included:

- Attendance for training was high, with 91% of participants (lead teachers) joining the training day face-to-face and 9% attending remotely. All lead teachers stated that the training was useful and pitched at the correct level.
- 92% of lead teachers participating in the second rollout reported that it was straightforward to share their learning with the wider team and that it was manageable to conduct the one session per week and deliver it in the way intended.⁶ Those who did not find it manageable reported this was due to summer term diary challenges, planned summative assessments, and strike days.

⁴ Different schools were involved in each round of the previous two rollouts. The schools who took part in the previous two rollouts will not be participating in this pilot.

⁵ EIAs are the third of local authorities in England where educational outcomes (for example, reading, writing, maths) were the weakest. This is based on sustained low performance across Key Stage 2 and Key Stage 4 in 2017 to 2019 (DfE, 2023a). As of March 2025, no policies are intended to be prioritised for EIAs.

⁶ The response rate was 100%. The early-stage programme development work came with a financial incentive for participating schools that was only paid on completion of all surveys, which facilitated completion rates.

- The problem-solving question bank and initial training were regarded as the most effective strands of support by participating teachers.

Some changes were made to programme delivery based on key learnings following the second rollout. These include:

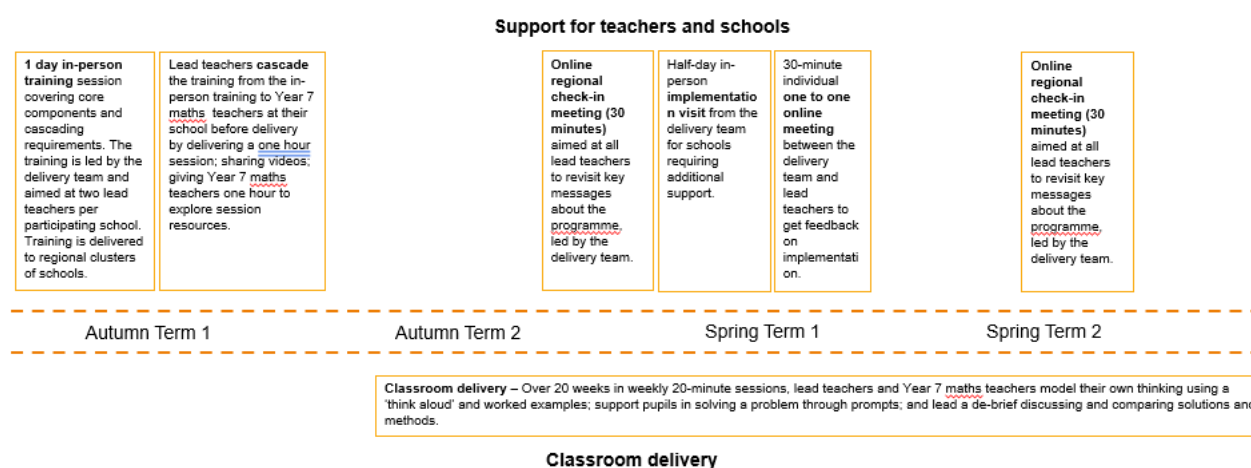
- **Intervention duration**—Participants in the second rollout suggested that a 10-week delivery period is too short to embed practice and achieve meaningful outcomes. Therefore, the length of the intervention was increased to 20 weeks.
- **Mode of ongoing support**—An element of the ongoing support provided to schools in the second rollout was an in-person visit to each school. This was replaced by an online meeting for each school in the pilot, as it would be logistically challenging for Blackpool Research School (BRS) to visit 30 schools in several regions where the programme is being delivered. However, in-person visits were planned for any schools that the delivery team identified as requiring additional support based on the responses provided in the delivery logs (see below).
- **Frequency of completing delivery logs**—Lead teachers completed the delivery logs during the second rollout on a weekly basis. The delivery logs are short forms (completed via Google Forms) covering the same questions each week on implementation. They are completed by participating lead teachers and shared with the delivery team. During the pilot, the delivery log was expected to be completed on a fortnightly basis instead to minimise the burden on schools.

The promising evidence, combined with the changes made to the programme, provide the context and rationale for evaluating the programme, with a specific focus on gathering evidence on the new programme elements.

Intervention

This section provides a detailed description of the intervention, with its components described based on a shortened version of the template for intervention description and replication (TIDieR) checklist (Humphrey *et al.*, 2016; see Appendix 1). Figure 1 provides a high-level overview of programme inputs described in this section.

Figure 1: Chronological overview of programme inputs



WHAT

Procedures

Initial training

The programme begins with two attendees (lead teachers) from each school attending a full-day face-to-face initial training session delivered by BRS. The training covers the key elements of the intervention. It also equips lead teachers with

resources to support implementation and cascading at their school, including sections on the delivery log, cascading expectations, and progress monitoring.

Cascading

The two lead teachers at each school are responsible for implementing a ‘cascading’ train-the-trainer (TTT) model at their school, following guidance set out by BRS. Minimum expectations for successful cascading include:

- Two lead teachers leading a cascading session for all Year 7 maths teachers before delivery begins. This session should be scheduled for at least one hour and be structured around the videos provided by BRS;
- Teachers having a minimum of one hour set aside before delivery starts to explore the lesson resources.

Alongside these requirements, BRS recommends that schools:

- Make time for teachers to rehearse the 'think aloud' approach in pairs or small groups before delivery (this is modelled during the face-to-face training day);
- Provide coaching support for teachers from one of the two lead teachers throughout delivery, including lesson drop ins and follow-on discussions (resources are provided by BRS to support this); and
- Have ongoing discussions about implementation by ensuring that programme delivery is a standing item on team meeting agendas for the duration of delivery.

Implementation

Delivery is overseen by one of the two lead teachers at each school (a ‘delivery lead’, who is usually, but not always, the head of the maths department).⁷ In each school, the approach is delivered by all Year 7 maths teachers once per week in their classes over a 20-week period. Delivered as part of usual maths lessons, each session takes 20 minutes, with flexibility to allow up to 30 minutes in the initial few weeks of implementation while the teachers get used to the programme.

During delivery, teachers use the question bank, handbook, and classroom resources provided by BRS to select appropriate problem-solving questions for use in lessons. In the weekly problem-solving sessions, teachers model their own thinking using a ‘think aloud’ method and appropriate worked examples. Teachers provide pupils with support using heuristic checklists and create debrief opportunities for pupils to discuss their methods and thought processes when solving mathematical problems. This might take the form of whole-class discussions led by teachers, pupils working out solutions in pairs, or individual reflection.

During implementation, the delivery lead in each school completes a delivery log on a fortnightly basis. The delivery log includes:

- Reflections on how the weeks went (what went well/less well);
- Whether the TM approach was implemented and if not, why not; and
- Duration and number of sessions.

The delivery logs are shared with BRS and used by BRS to identify schools who may benefit from additional implementation support (see below).

Ongoing support

Ongoing support is provided by BRS to all schools through (1) two regional online check-in meetings and (2) one-to-one online meeting. Schools requiring additional support receive (3) a supportive implementation visit.

Regional online check-in meetings – The online regional check-in meetings revisit key aspects of the teaching approaches, facilitate discussion between schools, and invite feedback. Each meeting lasts for 30 minutes and is attended by the same

⁷ The majority of delivery leads during the pilot were heads of their school’s maths department. In schools where this was not the case, this was because the head of maths did not teach a Year 7 class.

lead teachers who participated in the training sessions. The two check-in meetings are held during weeks 4/5 and weeks 12/13 of the 20-week programme.

Online meeting – A one-to-one online meeting is provided to all schools to allow the lead teachers to give any feedback on programme implementation that has not been captured in the delivery logs. Each meeting lasts for 30 minutes and is attended by the same lead teachers who participated in the initial training. They take place at around Week 10.

Implementation support visit – The delivery team organises specific in-person school visits for schools identified as needing additional implementation. Prior to delivery, BRS set out the criteria for support:

- A school fails to complete two delivery logs in a row despite reminders by BRS, or;
- After Week 3, the classroom sessions take longer than stipulated as per the delivery log i.e., longer than 20 minutes.

When selecting schools for implementation visits based on session duration, the delivery team looks for patterns of behaviour reported in the delivery log; that is, schools who *routinely* indicate that the sessions are taking longer than expected, rather than a one-off which can be attributed to other factors rather than implementation challenges, such as the complexity of specific problems. A further consideration when selecting schools who report that delivery takes longer than 20 minutes is to prioritise those schools who routinely report that some of their classes are taking longer than expected to deliver the sessions. This enables BRS to identify if certain types of groupings lead to implementation challenges for example, groupings of lower prior attainment pupils taking longer to complete the problems. During follow-up communication via email, the delivery team explained that they highlight to lead teachers during training and regional check-in meetings the importance of adhering to 20 minutes, given that participants might think ‘longer is better’.

The visits are carried out by a single team member from BRS and are aimed at addressing early-stage implementation challenges typically occurring within the initial eight weeks of the 20-week programme. The duration of each visit is tailored to the specific needs of the school, with the expectation that visits are around half a day long. During these visits, BRS observes implementation of the intervention in classrooms and provides tailored and constructive feedback. Lead teachers in each school facilitate the visit; however, the level of participation from other Year 7 maths teachers will vary by school and depend on the nature of the schools’ challenges.

Materials

Schools are provided with resources to support cascading and delivery following the in-person training sessions. These are provided all at once via a shared online drive and include:

- Video exemplification of expert teachers implementing TM in lessons;
- Programme handbook;
- Train-the-trainer (cascading) slides, guidance, and videos;
- Weekly problems for teachers⁸;
- Sample scripts to use with selected problems;
- Slides for teachers to use in lessons;
- Problem structure grid for use in lessons; and
- Heuristics for pupils to use in lessons.

HOW

The intervention is delivered in person to all Year 7 pupils during regular maths lessons. The training and ongoing support delivered to lead teachers by BRS includes a mix of in-person and online delivery. The initial one-day training is in person.

⁸ The protocol states that BRS would provide teachers with a problem bank to choose from at each weekly session. However, to ensure consistent delivery across all pilot schools and minimise the time teachers would spend choosing suitable problems, BRS pre-selected a weekly problem. This meant that all schools were required to use the same problem for each weekly session.

Ongoing support is both online (regional check-in meetings and one-to-one meetings) and in-person (implementation support visits).

Lead teachers implement the TTT model in-person at each participating school. They deliver cascading training sessions for all Year 7 maths teachers in participating schools. Additionally, all Year 7 maths teachers involved in the intervention are advised to meet in person once a fortnight for a minimum of 20 minutes to discuss implementation strategies, although this is optional.

WHERE

TM sessions are held during timetabled Year 7 maths lessons in the classroom.

WHO PROVIDED

The intervention has been created and is led by Simon Cox, Director of BRS and a secondary maths specialist. David Middleton works alongside Simon and is the Lead Teacher for the Research School.

In schools, the intervention is delivered by all Year 7 maths teachers who receive cascading training and support from the two lead teachers.

Research questions

The evaluation set out to answer a total of 15 research questions addressing the EEF’s three pilot criteria (see Table 2):

- Feasibility of implementation;
- Evidence of promise; and
- Readiness for trial.

The research questions were developed in consultation with the EEF and the delivery team during the project’s inception phase. The research questions about feasibility reflect the causal mechanisms and contextual assumptions set out in the ToC and associated logs. The research questions about evidence of promise address the intended and unintended consequences stated in the theory of change (ToC). The research questions about readiness for trial consider scalability and the ability of an evaluation to test the programme’s ToC, including intended impacts.

Table 2: Research questions

Pilot criteria	Research questions
Feasibility of implementation	<p>1. Fidelity: Are intervention activities (training, ongoing support, and delivery) implemented as intended?</p> <ul style="list-style-type: none"> • Are the two lead teachers attending training (trainers) able to successfully cascade the programme’s aim, resources, and processes with the Year 7 maths team (practitioners) in their school? • Does the Year 7 maths team effectively implement the programme’s approach in their teaching? Are practitioners more confident in using and applying maths resources?
	<p>2. Quality and dosage: Are intervention components of an appropriate quality?</p> <ul style="list-style-type: none"> • Are the resources and training materials provided useful for all schools? • Is the ongoing and bespoke support provided sufficient and effective? • Which resources or elements of the training and ongoing support do schools find most useful? • Does a 20-week model for delivery TM seem appropriate?
	<p>3. Acceptability: Are the resources, training, and support provided appropriate for schools with above-average numbers of FSM eligible pupils?</p>
	<p>4. Barriers and facilitators: What are the barriers and facilitators to successful implementation and take-up of the programme, including in schools with above-average numbers of FSM eligible pupils?</p>
	<p>5. Unintended consequences: Are there any unintended consequences of the programme’s delivery?</p>
Evidence of promise	<p>6. Teacher responsiveness: Is there evidence that teachers are motivated to engage with the programme and change their classroom practice?</p>
	<p>7. Cascading model: How effective is the train-the-trainer model in supporting programme objectives?</p> <ul style="list-style-type: none"> • Do the lead teachers feel confident in supporting the Year 7 teaching team? • Do practitioners feel knowledgeable about the programme approach (for example, the ‘talk aloud’ method, leading a debrief etc.)?
	<p>8. Teacher outcomes: Is there a change in teachers’ confidence and knowledge in teaching tackling of unfamiliar mathematical problems?</p> <ul style="list-style-type: none"> • Do lead teachers and practitioners have improved understanding of how to adapt teaching to explicitly address problem-solving challenges? • Do lead teachers and practitioners feel confident in using the TM approach and delivery in lessons? • Do lead teachers and practitioners more regularly use appropriate mathematical problems in lessons, ‘talk aloud’ their processes for solving mathematical problems, and incorporate time for discussing different problem-solving strategies into lessons?

	<p>9. Pupil responsiveness: Is there evidence that pupils are motivated to engage in the programme?</p>
	<p>10. Pupil outcomes: Is there a (perceived) change in pupils' approaches to unfamiliar mathematical problems?</p> <ul style="list-style-type: none"> • Is there a (perceived) change in how pupils deal with mathematical problems (for example, use of heuristics)? • Is there a (perceived) change in pupil knowledge, confidence, and enjoyment of mathematical problem-solving? • Is any (perceived) change observed for all pupils, including those from disadvantaged backgrounds?
	<p>11. Sustainability: Does the programme appear to be sustainable?</p> <ul style="list-style-type: none"> • Are changes made by schools and teachers at the start of the programme and/or behaviour changes sustained over the 20-week implementation period? • Do school staff perceive that these changes will continue to be sustained after the 20-week implementation period?
Readiness for trial	<p>12. Codification: Is the programme sufficiently manualised to allow delivery with fidelity to a larger number of schools?</p> <ul style="list-style-type: none"> • Is the codification for the ongoing support to schools appropriate?
	<p>13. Monitoring: Are viable strategies in place to collect sufficient data to monitor compliance and fidelity?</p>
	<p>14. Outcome measurement: What outcome measures might be used in an efficacy trial?</p>
	<p>15. Scalability: Is the delivery approach appropriate to be delivered with fidelity at larger scale?</p> <ul style="list-style-type: none"> • Does the delivery team have sufficient capacity to deliver the programme with fidelity at larger scale?

In this report, we present evidence against 18 key success indicators to inform the decision about whether TM should proceed to trial. The success indicators were developed in close collaboration with the delivery team (DT) and the EEF during the inception phase to ensure agreement on what constitutes 'success' for each research question. Table 3 maps these success indicators against the three pilot criteria and the research questions listed above and provides the sources of evidence for each indicator and how they will be assessed. The sources of evidence are all explained in more detail in the Methods section.

At the end of each chapter, we present a short summary of the success indicators and summarise the evidence against each. At the end of the report, we then present an assessment for each indicator, commenting on the extent to which they have been successful or not.

Table 3: Pilot criteria

Pilot criteria	Success indicators	Sources of evidence
Feasibility of implementation	Programme activities (training/ongoing support/in-class delivery) are delivered with fidelity (RQ1)	<ul style="list-style-type: none"> • Programmatic data • Interviews with DT • Pre-post teacher survey • Paired interviews with lead teachers • Focus groups with practitioners • Train-the-trainer (TTT) training observations and ongoing support • Classroom delivery observation

	Intervention components (training/ongoing support/training material/delivery resources) are of appropriate quality (RQ2)	<ul style="list-style-type: none"> • Programme documentation • Interviews with DT • Pre-post teacher survey • Paired interviews with lead teachers • Focus groups with practitioners • Paired interviews with pupils • TTT training observations and ongoing support • Classroom delivery observation
	Resources/training/ongoing support is appropriate for schools with above-average proportion of FSM pupils (RQ3)	<ul style="list-style-type: none"> • Interviews with DT • Learning workshop • Paired interviews with lead teachers • Focus groups with practitioners
	Barriers to delivery do not impede the programme being delivered with fidelity (RQ4)	<ul style="list-style-type: none"> • Interviews with DT • Pre-post teacher survey • Paired interviews with lead teachers • Focus groups with practitioners • Paired interviews with pupils • TTT training observations and ongoing support • Classroom delivery observation
Evidence of promise	Teachers are motivated to engage in the programme and change classroom practice (RQ6)	<ul style="list-style-type: none"> • Interviews with DT • Pre-post teacher survey • Paired interviews with lead teachers • Focus groups with practitioners • TTT training observations and ongoing support
	<p>The train-the-trainer approach results in lead trainers feeling confident in supporting the Year 7 teaching team (RQ7)</p> <p>The train-the-trainer approach results in practitioners feeling knowledgeable about key elements of the programme (RQ7)</p>	<ul style="list-style-type: none"> • Pre-post teacher survey • Paired interviews with lead teachers • Focus groups with practitioners • Classroom delivery observation
	Teachers are more knowledgeable and confident in teaching the tackling of unfamiliar mathematical problems (RQ8)	<ul style="list-style-type: none"> • Interviews with DT • Pre-post teacher survey • Paired interviews with lead teachers • Focus groups with practitioners
	Pupils are motivated to engage in the programme (RQ9)	<ul style="list-style-type: none"> • Paired interviews with lead teachers • Focus groups with practitioners • Paired interviews with pupils • TTT training observations and ongoing support
	<p>Positive change in pupils' approaches to tackling unfamiliar mathematical problems (RQ10)</p> <p>No evidence that disadvantaged pupils benefit less than their less disadvantaged peers (RQ10)</p>	<ul style="list-style-type: none"> • Pre-post teacher survey • Paired interviews with lead teachers • Focus groups with practitioners • Paired interviews with pupils
	Changed practice by school and teachers embedded throughout 20 weeks and intended to be sustained after 20 weeks (RQ11)	<ul style="list-style-type: none"> • Programmatic data • Pre-post teacher survey • Paired interviews with lead teachers • Focus groups with practitioners
Readiness for trial	Core components identified for key programme activities (RQ12)	<ul style="list-style-type: none"> • Programme documentation • IDEA workshop • Interviews with DT • Learning workshop

<p>Sufficient data collected to monitor compliance and fidelity (RQ13)</p> <p>Time spent on data entry (for example, completing delivery logs) is proportionate to its purpose (RQ13)</p>	<ul style="list-style-type: none"> • Programme documentation • Programmatic data • IDEA workshop • Interviews with DT • Learning workshop • Pre-post teacher survey
<p>Outcome measure(s) identified that is/are:</p> <p>a) Aligned to intended outcomes;</p> <p>b) Able to measure changes that we expect to see; and</p> <p>c) Convenient to administer in secondary schools (RQ14)</p>	<ul style="list-style-type: none"> • Programme documentation • Outcome measure review
<p>Clear and realistic plan for what scaling to deliver in more schools would look like and require (RQ15)</p> <p>Plan for scaling would entail little or no modification to content of in-school delivery (RQ15)</p>	<ul style="list-style-type: none"> • Interviews with DT • Learning workshop

Ethical review

The project was reviewed and approved by the Oxford MeasurEd ethics board in April 2024. Two external, independent reviewers reviewed the pilot plan and an ethics application outlining processes and protocols regarding recruitment and consent, protection from harm, incentives and recognising participation, confidentiality and data sharing, safeguarding, accessibility, inclusion and anti-racism, environmental impact, and other ethical issues. They provided their comments to the board Chair, a Director of Oxford MeasurEd who was not involved in the delivery of the project. The Chair then shared the comments with the project team, and the project team agreed on actions to be taken forward. The Project Director approved these actions, and the board Chair signed off on the approval.

All of Oxford MeasurEd’s evaluation work is guided by UK Evaluation Society Guidelines for Good Practice. The evaluation team was required to comply with our Code of Conduct, Safeguarding, and Anti-Bribery and Corruption policies.

School agreement was obtained through a Memorandum of Understanding (MoU) (Appendix 2). The MoU was signed by a member of the senior leadership team (SLT) managing the maths department as well as both participating lead teachers and sent to the delivery team. The delivery team then shared the signed MoUs with the evaluation team via a shared folder so that the evaluation team had the schools’ and lead teachers’ contact details.

Data protection

Data collected for this evaluation was only used for research purposes. We stored and handled data in line with the General Data Protection Regulation (GDPR) and the Data Protection Act 2018. Only named individuals had access to personal data, and the team complied with information security procedures that include preventative measures and processes for reporting, reviewing, and responding to breaches. All data transferred to Oxford MeasurEd for the purposes of the evaluation was done via secure cloud software, and Oxford MeasurEd did not transfer the data to any other party. We will securely delete personal data six months after the project ends.

We pseudonymised all data, removing schools’ and individuals’ names prior to analysis. We did not report schools’ or individuals’ names in evaluation outputs, but some participants may be recognisable due to the small size of case study schools. We also communicated these limits to confidentiality and anonymity verbally to all interview participants before the interviews. We also checked in with participants after the interviews to ensure that they were happy for everything they said to be possibly included in the evaluation report.

For this evaluation, Oxford MeasurEd was a data controller who also processed data. This means that we are responsible for deciding the purpose and legal basis for processing data. Our legal basis for data processing was ‘legitimate interest’. Data subjects were the delivery team, lead teachers, practitioners, and pupils. We also outlined our data

protection procedures and safeguards and our legal bases for processing data during recruitment and published a privacy notice online, which was circulated to all concerned parties.

All evaluation data will be securely deleted from Oxford MeasurEd's systems no more than six months after submission of the final edited EEF report (by June 2026).

Project team

The project was delivered by a team from Blackpool Research School (BRS), based at St Mary's Catholic Academy in Blackpool.⁹

Table 4: Delivery team

Name	Role	Institutional affiliation
Simon Cox	Director of Research School	BRS
David Middleton	Lead Teacher	BRS

The evaluation was delivered by the following team from Oxford MeasurEd.

Table 5: Evaluation team

Name	Role	Institutional affiliation
Dr Lydia Marshall	Principal Investigator and Project Director	Oxford MeasurEd (Director of Research)
Dr Jonah Bury	Senior Evaluator and Project Manager	Oxford MeasurEd (Principal Consultant)
Robert Wishart	Impact Evaluation Expert	Oxford MeasurEd (Associate)
Sara Bashir Malik	Junior Evaluator and Project Coordinator	Oxford MeasurEd (Senior Consultant)
Ellen Smith	Junior Evaluator	Oxford MeasurEd (Consultant)

Methods

Design

In line with the EEF's Implementation and Process Evaluation (IPE) guidance (EEF, 2022a) and pilot guidance (EEF, 2023), we adopted a mixed-methods approach to answering the research questions. This was important to ensure the robustness of findings as well as reducing the limitations of relying solely on qualitative or quantitative methods. Our design included desk-based research, primary research with the delivery team, lead teachers, Year 7 teachers, and pupils, and observations of TM delivery. Our design is further characterised by:

- Focusing on those elements that are 'new' to the programme or where evidence gaps remain;
- Focusing on short-term (proximal) outcomes to demonstrate evidence of promise;
- Conducting intensive research with a smaller number of schools, enabling us to a) triangulate data sources to reach a deeper understanding of implementation and b) better incentive participation through larger financial incentives for fewer schools;
- Capturing pupils' voices through a paired discussion group, enabling us to hear from pupils directly about their views of the programme and any benefits they might have experienced; and
- Leveraging programme monitoring data collected by the delivery team to assess fidelity.

⁹ Initially, the delivery team included three members of staff (see Marshall *et al.*, 2024); however, BRS decided following the recruitment stage that it was feasible to deliver the pilot with two members of staff.

Recruitment

BRS recruited 30 secondary schools for the pilot evaluation.¹⁰ We agreed the sample size with the EEF and BRS, considering both BRS's delivery capacity and the sample size required to enable exploration on questions of scalability and fidelity. During the recruitment period between April and June 2024, BRS, EEF, and Oxford MeasurEd monitored recruitment to ensure that:

- Participating schools were not participating in other EEF-funded maths/numeracy-focused interventions in Year 7, to avoid contamination;
- The sample included a mix of schools from urban and rural areas;
- There was a broadly even split between the two participating regions (North West and Yorkshire and the Humber); and
- Around one-third of schools were located in Education Investment Areas (EIAs).

The aim to include a broadly even split between the two participating regions was dismissed during recruitment, as there was more interest from schools in the North West region. As part of the signing up process, participating schools were asked to nominate two lead teachers. Both lead teachers needed to be subject specialists and one lead teacher the school's head of maths. Beyond this, there were no eligibility criteria for lead teachers, other than the school authorising their time to be spent attending training, cascading the training to other Year 7 maths teachers at their school, and delivering the intervention. The lead teachers and a member of the senior leadership team (SLT) signed the Memorandum of Understanding (MOU) agreeing to the requirements of the programme and the evaluation. The MOUs were shared by BRS with the evaluation team via a secure platform (Google Drive) on a regular basis during the recruitment period.

The evaluation team prepared all recruitment materials for evaluation activities and shared them with BRS and the EEF for review (see Appendices 2, 3, 4, and 5). These recruitment materials outlined the purpose of the evaluation, what would be involved in taking part, and how data would be used. We also explained this verbally before all workshops, interviews, and observations, and participants were able to ask questions before agreeing to take part. Recruitment documents and verbal briefings also explained the steps we would take to minimise the chance of participants being identifiable in reports and explained how people could let us know if there was anything they would not want us to include in the report.

Data collection

Our design included desk-based research, primary research with the delivery team, lead teachers, Year 7 teachers, and pupils, and observations of TM delivery.

Figure 2 sets out how our design addressed the three pilot criteria and 15 research questions.

¹⁰ Initially the aim was to recruit 35 schools. Slight challenges with recruitment meant that the target number of schools was reduced to 30 at the end of June 2024.

Figure 2: Evaluation matrix

Evaluation domain	Evaluation questions	Desk-based research			Research with DT			Research with teachers			Research with pupils	Observations	
		Programme documentation	Programmatic data	Outcome measure review	IDEA workshop	Interviews with DT	Learning workshops	Pre-post teacher survey	Paired interviews with lead teachers	Focus groups with practitioners	Paired interviews with pupils	TTT training and ongoing support	Classroom delivery
Feasibility	1. Fidelity: Are intervention activities (training, ongoing support, and delivery) implemented as intended?												
	2. Quality and dosage: Are intervention components of an appropriate quality?												
	3. Acceptability: Are the resources, training, and support provided appropriate for schools with above-average numbers of FSM eligible pupils?												
	4. Barriers and facilitators: What are the barriers and facilitators to successful implementation and take-up of the programme, including in schools with above-average numbers of FSM eligible pupils?												
	5. Unintended consequences: Are there any unintended consequences of the programme's delivery?												
Evidence of promise	6. Teacher responsiveness: Is there evidence that teachers are motivated to engage in the programme and change their classroom practice?												
	7. Cascading model: How effective is the train-the-trainer model in supporting programme objectives?												
	8. Teacher outcomes: Is there a change in teachers' confidence and knowledge in teaching tackling of unfamiliar mathematical problems?												
	9. Pupil responsiveness: Is there evidence that pupils are motivated to engage in the programme?												
	10. Pupil outcomes: Is there a (perceived) change in pupils' approaches to unfamiliar mathematical problems?												
	11. Sustainability: Does the programme appear to be sustainable?												
Readiness for trial	12. Codification: Is the programme sufficiently manualised to allow delivery with fidelity to a larger number of schools?												
	13. Monitoring: Are viable strategies in place to collect sufficient data to monitor compliance and fidelity?												
	14. Outcome measurement: What outcome measures might be used in an efficacy trial?												
	15. Scalability: Is the delivery approach appropriate to be delivered with fidelity at larger scale?												

Workshop, surveys, interviews, and focus groups were carried out and/or led by the evaluation team. The delivery team collected the programmatic data in the form of attendance data and the delivery logs every two weeks, which the delivery team shared with the evaluation team at the end of delivery. To encourage open responses, the delivery team provided assurances that any data shared with the evaluation team would be anonymised.

Table 6 provides an overview of the different participant groups; the activity, mode, and duration of the research encounter; as well as the overall number of research encounters and/or the overall sample size and the target number in brackets. Case study schools received a thank you payment in the form of a £200 voucher for facilitating the visits.

Table 6: Overview of participant groups and activities

Participant group	Activity	Mode and duration	N (target)
Delivery team	IDEA workshop	In person, 120 mins	1 workshop
	Qualitative interviews	Online, 45 mins	4 interviews (6)
	Learning workshop	In person, 120 mins	1 workshop
Teachers	Pre-delivery survey	Online, c. 10 mins	82 Year 7 teachers (140) 47 lead teachers (56)
	Post-delivery survey	Online, c. 15 mins	51 Year 7 teachers (140) 34 lead teachers (56)
	Paired interviews with lead teachers	In person or online, 45 mins	10 (10)
	Focus group discussions	In person or online, 60 mins	8 (10) focus group discussions including 2–4 teachers per group
Pupils	Paired interviews	In person or online, 30–45 mins	10 interviews (10)
Observations	Initial training session	In person, all day	2 all day training sessions (2)
	Regional check-in meetings	Online, 30–45 mins	2 (4)
	In-school delivery	In person, c. 20 mins	10 sessions (10)

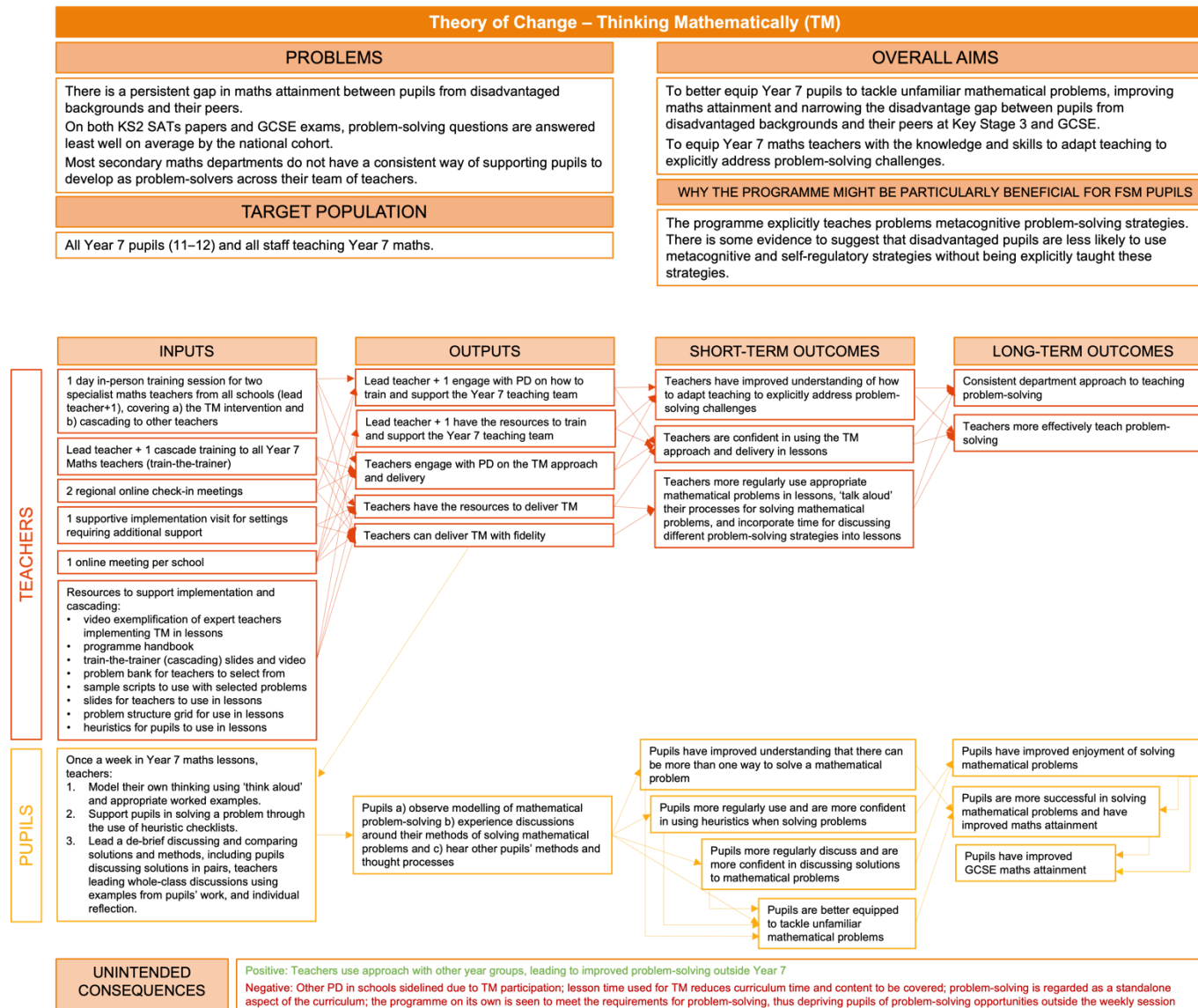
Desk-based research

Before finalising the evaluation design, we **reviewed programme documentation** to:

- Understand the ToC (Figure 3), including causal mechanisms and contextual assumptions;
- Map research evidence collected against the causal mechanisms and contextual assumptions so far, and identify evidence gaps; and
- Identify any programmatic data that we might use for the evaluation.

Ahead of the first set-up meeting in February 2024, we presented an updated ToC to EEF and the delivery team. The updated ToC built on the ToC that the delivery team had developed with the EEF during workshops before the pilot. A new addition compared to the original ToC included a box on potential positive and negative unintended consequences that schools, teachers, and pupils could experience as a result of programme participation. Following the set-up meeting, the ToC was refined iteratively with input from the delivery team and finalised at the end of February 2024.

Figure 3: Theory of Change (developed by the delivery team with facilitation from the EEF and the evaluation team)



Once we received a full set of programme resources, we appraised a) the *quality* of resources (for example, problem bank) and b) the level of *codification*. We operationalised these concepts via the following questions:

- Is it clear what the teacher needs to do step-by-step?
- Are the instructions documented and accessible (for example, non-technical language)?
- Are roles, timings, and materials specified?
- Are all components defined?
- Is anything missing that requires tacit knowledge?

We triangulated our appraisals with feedback received from research participants to analyse resource quality and accessibility.

We carried out a desk-based **review of potential outcome measures** for a future trial. We agreed the methodology for this review with the EEF. Our review included considering, for each potential outcome measure:

- Whether they were suitable for testing Year 7 maths problem-solving;
- Their statistical properties; and
- Logistical factors such as mode of data collection, length of assessment, and any training/briefing required to administer the assessment.

We analysed the pros and cons of each identified measure and outlined these for discussion with the EEF and BRS. We identified the Progress Test in Maths published by GL Assessments as an appropriate measure for testing Year 7 maths pupils' problem-solving skills. The Progress Test in Maths is a comprehensive assessment of attainment administered orally to each child individually. This measure was chosen because it particularly targets problem-solving skills, has been designed for use in the UK, and has been UK norms tested using a large sample size. Moreover, it has been successfully used in previous EEF trials, can be administered to the whole class digitally or as a paper-based assessment, and has good construct validity and internal consistency. With the EEF and BRS, we collectively decided against piloting the tool. A note outlining the methodology and the reasons for our decision can be found in Appendix 12.

Finally, we collected **programmatic data** to understand how TM was delivered and assess fidelity and feasibility. This included quantitative measures of dosage (attendance at training and ongoing support, completion of delivery logs), and the quantitative content of the delivery logs, including the length of each weekly session, the extent of use of programme materials, and their perceived usefulness.

Research with the delivery team

We facilitated an **IDEA workshop** with the delivery team in January 2024 to discuss the ToC, assumed causal mechanisms, and contextual assumptions. Following the workshop, we finalised the ToC in close consultation with the delivery team and the EEF.

During implementation, we carried out four **individual interviews**¹¹ with the two members of the delivery team at two separate timepoints: in December 2024 and March 2025.¹² The first interview gathered views on the training session and the feasibility of cascading and delivering ongoing support. The second round of interviews focused on barriers and facilitators to implementation, reflections on evidence of promise in terms of teacher- and school-level change as well as initial ideas around scaling. Each interview lasted between 45 and 60 minutes and took place remotely.

We also facilitated a two-hour in-person post-delivery learning workshop with the delivery team in July 2025¹³ to explore lessons learned by the team, discuss the delivery team's plans for scaling delivery, and seek participant validation of our interim findings.

¹¹ Because the delivery team only consisted of two members rather than three, we did not carry out six interviews as originally planned.

¹² The study plan states that the interviews with the delivery team would take place in November 2024 and February 2025. However, interviews with the delivery team took place a bit later to ensure they happened after the regional check-in meetings.

¹³ The learning workshop was not held in August 2025 as initially planned due to the school holidays.

Research with school staff

Qualitative research with school staff

Between January and April 2025, we carried out qualitative data collection with school staff at case study schools. The aim of the **(paired) interviews with lead teachers** and **focus groups with Year 7 teachers** was to explore questions related to evidence of promise, feasibility, and readiness for trial. The overall aim was to carry out ten paired interviews with lead teachers. The rationale was to interview *both* lead teachers in a school simultaneously to ensure a more in-depth understanding of day-to-day implementation. We also aimed to conduct ten focus groups with Year 7 teachers. We invited all lead teachers at the case study schools to take part in the interviews as well as inviting those Year 7 teachers who were available for a focus group discussion on the day of our visit. We did not stipulate a required number of Year 7 teachers for the focus group, as we were conscious of minimising burden on schools and did not expect all participating Year 7 teachers to be available for a focus group discussion at the same time.

Overall, we completed ten interviews with lead teachers, of which four interviews were paired. We completed eight focus groups with Year 7 teachers, two less than the overall target. During two case study visits, we were unable to carry out focus groups with Year 7 teachers due to competing priorities and staff absence on the day of our visit. The focus groups ranged from two to four participants per group.

All interviews and focus groups were held in person during our case study visits, except for one lead teacher interview held online, as they were unavailable for an interview on the day of our visit. Before each data encounter, we underlined that only the research team had access to the data and provided reassurances around anonymity. At the end of each encounter, we checked in with participants to ensure they were happy for everything that they discussed to be included in the report. Interviews lasted around 60 minutes, although there was some variation in interview length for example, due to participants' other commitments. We facilitated interviews and focus groups using a topic guide, designed to allow for flexibility in conversations while ensuring that key topics and themes were covered (see topic guide example in Appendix 7). To ensure consistency in data collection and minimise bias, a senior member of the evaluation team quality assured the first interview and focus group carried out by each team member.

Quantitative surveys of school staff

We invited all lead teachers and Year 7 teachers to complete a short online survey pre-delivery in September and October 2024 to capture teachers' baseline awareness of the components targeted by the programme, their confidence and understanding in teaching problem-solving, and their perceptions of available resources and teaching practices. The pre-delivery survey had 23 items. The data was collected from lead teachers before they received any training from BRS and for Year 7 teachers before the training was cascaded. This was to ensure we captured their understanding, skills, confidence, and perceptions *before* they could have been influenced by the programme. The purpose of this was to ensure we captured the true picture of teacher perceptions. Overall, 47 out of 55 lead teachers (85%) and 82 out of 139 Year 7 teachers (59%) successfully completed the pre-delivery survey. The average completion time for the pre-delivery survey was approximately ten minutes.

Following programme completion, we invited lead teachers and Year 7 teachers separately to complete an online post-delivery survey between late April and early July 2025¹⁴ to quantify their experiences with implementation, the quality and relevance of training and resources, perceived barriers and facilitators, perceptions of benefits and unintended consequences, and intentions to continue using the programme practices. The survey also collected data on their confidence and understanding of teaching problem-solving, and their perceptions of available resources and teaching practices following programme delivery. The post-delivery survey had 56 items for lead teachers and 49 for Year 7 teachers. Overall, 34 out of 55 lead teachers (62%) and 51 out of 142¹⁵ Year 7 teachers (36%) successfully completed the post-delivery survey. A lower response rate among Year 7 teachers may reflect lower perceived ownership of the programme compared

¹⁴ The time window for completing the post-delivery survey was extended several times due to the low response rate. We also asked the EEF to send a letter to schools to encourage participation.

¹⁵ More Year 7 teachers received the post-delivery survey than the pre-delivery survey because additional email contacts were identified during delivery, including newly appointed teachers and others who had not been included earlier.

with lead teachers, who were more directly responsible for its implementation. Of the post-delivery survey respondents, 29 out of 34 lead teachers (85%) and 33 out of 51 Year 7 teachers (65%) had also completed the pre-delivery survey. The average completion time of the post-delivery survey was approximately 13 minutes for both lead teachers and Year 7 teachers.

Observations of training, ongoing support, and in-school delivery

We observed two initial training sessions in September 2024 in two different regions to understand any differences in delivery. We observed both regional check-in meetings in December 2024 and March 2025.¹⁶ We did not attend the meetings live but observed them asynchronously due to time constraints and to minimise disruption to the sessions. Finally, we observed ten mathematical problem-solving sessions with pupils—one during each of our case study visits.

The primary aim of the observations was to provide insights to tailor prompts and probes used in interviews and focus groups. However, we also gathered observational data on fidelity of implementation, teachers' and pupils' responsiveness, quality of training and instruction, and any challenges to successful delivery.

Research with pupils

We carried out ten paired interviews with pupils as part of our case study visits between January and April 2025. They took place *after* our observation of the weekly sessions, were supervised by a familiar adult for safeguarding purposes, and lasted around 25 minutes. The adult was either the Year 7 maths teacher, another teacher, or school staff. The purpose of the interview was to explore pupils' ability to talk through a mathematical problem they had recently addressed and gauge their engagement,¹⁷ explore whether they enjoyed the programme, and investigate any barriers they experienced as well as any perceived outcomes from the programme.

Before our visit to each school, we instructed the lead teacher to select two pupils for the paired interview. We specified that we wanted them to select pupils who had not opted out of the study (either themselves or via their parents/carers), were in the same class (ideally the one observed during our visit), and were friends or comfortable being interviewed together. We also noted that it would be helpful for the pupils to have different levels of prior attainment in maths to ensure we included different ability levels. As far as we could determine, teachers adhered to our selection criteria; however, it was difficult to assess the extent to which they selected pupils with different ability levels based on a short paired interview (see Limitations).

Our approach to interviewing the pupils drew on good practice guidance (see for example, Brady and Graham, 2019). Specifically, this meant keeping activities brief, being flexible, and using age-appropriate and tailored language (see Appendix 7). For instance, we did not refer to the programme as TM but used the language that pupils were familiar with to describe what we wanted to talk to them about. When asking about outcomes, we asked the pupils to complete a ranking exercise by listing all the benefits on a piece of paper and then selecting the most important benefit. This enabled us to ask more targeted questions and gather rich data. After drafting the topic guide and the activities, we sought feedback from BRS as experts in both the intervention and in working with this age group.

Data analysis

We recorded workshops, interviews, and focus groups with participants' permission, using the 'Framework' approach to managing qualitative data (Ritchie *et al.*, 2013). Using the themes covered in topic guides, we assembled a matrix in which each row represents an individual interview and each column a theme, and summarised data in the matrix, including illustrative quotes where appropriate. Once we coded all data in this matrix, we analysed it thematically. This included applying the themes from topic guides that the matrix is structured by and identifying new themes emerging from the data,

¹⁶ The delivery team provided one regional check-in per timepoint; the study plan states that two regional check-in meetings would be provided at each timepoint. As a result, we only observed two check-in meetings across two timepoints and not four.

¹⁷ The initial plan had been to ask pupils to share their approach through a visualiser, which is a piece of technology used in classrooms and similar to an overhead projector; however, in practice, this was not feasible to do as the interviews were held in rooms without access to a visualiser.

meaning that the analysis was in part deductive and in part inductive. This enabled us to carry out both descriptive and explanatory analysis, identifying convergence and dissonance between participant groups, and looking for explanations for differing views.

We analysed quantitative survey and programmatic data descriptively using unweighted frequencies and ‘crosstab’ analysis in Stata. We agreed on a detailed analysis plan with the EEF. The Principal Investigator (PI) quality assured the analysis syntax and output, including confirming consistency with the analysis plan and carrying out spot checks to check for any errors or omissions in the analysis. Since there was a low response rate to the post-delivery survey, particularly for Year 7 teachers, the findings should be interpreted cautiously, as results may be subject to non-response bias and may not fully represent the experiences or views of all Year 7 teachers involved in delivery (see Limitations). As this was a pilot study, we prioritised reporting descriptive findings over hypothesis testing. However, for crosstabulations, we carried out tests of statistical significance and presented *p*-values only for differences that were statistically significant. Given the relatively small sample sizes, any statistically significant findings reflect differences large enough to be detectable despite limited statistical power. Overall, the hypothesis testing should be interpreted as exploratory, and the results viewed as indicative rather than definitive. While *p*-values have been presented in the report and appendices for transparency, they are not considered central to decision-making at the pilot stage. Instead, results should be interpreted holistically, alongside process data and evidence of readiness for a full trial (see Limitations).

Our analysis was ongoing throughout the pilot study. This ensured that we could provide formative outputs and follow-up on emerging findings in later research activities wherever possible. Throughout, we triangulated quantitative and qualitative findings to address the three evaluation domains and final evaluation questions. This triangulation culminated in an internal analysis workshop led by the PI. Here, the evaluation team came together to discuss qualitative and quantitative findings from different sources under each of the research questions and identify convergent and divergent findings. This form of peer discussion and review for mixed-method research also served the purpose of quality assurance, as it ensures the legitimacy of findings. Our reporting process also included participant validation at a learning workshop with the delivery team. At the learning workshop, we presented the emerging findings and sought reflections from the delivery team to contextualise and enrich our final conclusions and recommendations. Table 7 includes a timeline of activities related to the evaluation and intervention delivery including the recruitment period, data collection, and delivery schedule.

Timeline

Table 7 provides an overview of the timeline, including the date and the relevant programme and/or evaluation activity.

Table 7: Timeline

Date	Activity
Jan 2024	IDEA workshop with delivery partners
Feb 2024	Set-up meetings
Feb – Mar 2024	Desk-based review of outcome measures
Feb – Mar 2024	Recruitment materials
Mar – Apr 2024	Desk-based review of programme documentation
Mar – Apr 2024	Ethical approval
Mar – May 2024	Study plan
Apr – Jun 2024	School recruitment
Jun – Aug 2024	Instrument development
Jul – Sept 2024	School and teacher enumeration
Sept 2024	Initial training

Sept 2024	Training observations
Sept 2024	Cascading
Sept – Oct 2024	Pre-delivery teacher survey
Sept 2024 – Apr 2025	In-school delivery Ongoing support to schools from BRS
Dec 2024	Ongoing support observations Delivery team staff interviews Analysis of programmatic data
Jan – Mar 2025	Lead teacher interviews Year 7 teacher focus groups In-school classroom observations Pupil paired interviews
Mar 2025	Ongoing support observations Delivery team staff interviews
May – Jul 2025	Post-delivery teacher survey
Jun 2025	Analysis of programmatic data
Jul 2025	Interim findings brief
Jul 2025	Learning workshop with delivery team
Sept – Oct 2025	Final analysis of main evaluation data
Oct – Nov 2025	Final report and presentation

Findings

In this section, we begin by outlining the pilot participants and the final sample achieved across the evaluation activities. We then present the findings structured around the three evaluation domains: feasibility of implementation, evidence of promise, and readiness for trial. The research questions are addressed in chronological order. At the end of each section, we summarise the relevant success indicators and provide an overview of the evidence of success.

Participants

Thirty secondary schools from two regions in England signed up for the TM pilot, with 28 schools from the North West and two schools from the Yorkshire and Humber region. One school withdrew before programme delivery started, and another school did not take part in the evaluation to avoid any conflict of interest, as this was where the two key members of BRS were employed as teachers. The remaining 28 schools stayed in the evaluation throughout the pilot. Of those, 11 schools were situated in rural areas and 17 schools in urban areas.

The following section provides a short overview of the areas, schools, pupils, and practitioners involved in the pilot.

Area overview

Many areas in the North West region fall among those with the lowest household incomes in England (Office for National Statistics, 2021). The North West region has both the second highest rate of poverty (Joseph Rowntree Foundation, 2025) and the highest rate of child poverty in the country (Action for Children, 2024). Between 25% and 30% of working-age adults in the region are not in employment, compared to around 20% in regions with the lowest levels of poverty (Joseph Rowntree Foundation, 2025), with the unemployment rate twice as high for ethnically diverse people compared to the overall population (Business in the Community, 2023a). Ethnically, 14% of the population is from diverse backgrounds, with 86% identifying as White (Business in the Community, 2023a).

Similarly, many areas in Yorkshire and the Humber feature among England's lowest household income areas (Office for National Statistics, 2021). Yorkshire and the Humber has the fourth highest rate of poverty and child poverty in the country (Joseph Rowntree Foundation, 2025; Action for Children, 2024). The region has a lower life expectancy than the national average, ranking third lowest for life expectancy compared with other regions (Office for Health Improvement and Disparities, 2021). The majority (85%) of Yorkshire and the Humber's population are White. Among the non-White population, Pakistanis comprise the largest ethnic minority group in Yorkshire and the Humber (Business in the Community, 2023b).

School and pupil profile

Different school types were represented in the pilot. Nineteen (68%) of the 28 schools were academies, of which 16 (57%) belonged to multi-academy trusts (MATs), and three (11%) were single academies. Moreover, nine pilot schools (32%) were local authority (LA) maintained schools. The pilot sample was therefore not fully representative of the national picture, where 80% of secondary schools in England are academies (Department for Education, 2023, 2025).

There was some variation in the pilot schools' Ofsted¹⁸ ratings, although the majority ($n = 18$) (64%) of the participating pilot schools were rated 'good' by Ofsted. Five schools (18%) were rated 'requires improvement', four schools (14%) were rated 'outstanding', and one school (4%) was rated 'inadequate'. This broadly reflects the national picture of school inspection ratings at the time of the pilot (Education Policy Institute, 2024).

Schools grouped their Year 7 maths pupils in different ways. Eleven of the pilot schools (39%) grouped pupils in classes set by prior attainment (including both maths and overall attainment), ten schools (36%) grouped pupils through a combination

¹⁸ From November 2025, a new Ofsted framework will come into effect, which will remove single-word overall judgements with a new 'report card' system.

of mixed-ability classes and classes set by attainment, six schools (21%) grouped pupils by mixed-ability classes only, and one school (4%) conducted assessments at the beginning of Year 7 in order to set their pupils.

Schools ranged in size¹⁹, with the smallest school having a total enrolment figure of 367 and the largest school with a total enrolment figure of 2196. Twelve of the schools (43%) had above-average enrolment, while 16 of the schools (57%) had below-average enrolment (Homes England, 2023). As a result, there was variation in the number of Year 7 classes that took part in the pilot at each school.

Schools had varied proportions of pupils eligible for FSM, ranging from 9% to 63%. Sixteen of the pilot schools (57%) had an above-average proportion of students eligible for FSM (Department for Education, 2025), while ten of the 28 participating schools were located in EIAs. Though we did not collect data on the ethnic makeup of the schools, some lead teachers and teachers noted in interviews and focus groups that their schools were characterised by high levels of ethnic diversity, while others stated that their school was predominantly White.

Achieved sample

Research with school staff

Quantitative surveys of school staff

Pre-delivery survey

Lead teachers

The response rate for lead teachers was 85% (47 out of 55). The lead teacher sample at pre-delivery showed a reasonable age spread (see Table 8). The spread of teaching experience was skewed towards more experienced practitioners, with over half (52%) having more than 10 years of experience, while very few were early-career teachers. Subject specialism was also highly concentrated, with the majority of lead teachers being specialist maths teachers (91%). However, these levels of teaching experience and subject specialism were to be expected given the lead teachers' leadership roles.

Year 7 teachers

The response rate for Year 7 teachers was lower compared to lead teachers. Overall, 59% (82 out of 139) responded to the pre-delivery survey. The Year 7 teacher sample showed a fairly even spread across age groups, with around one-third (32%) aged 31–40 and another third (32%) aged 41–50. Teaching experience was more mixed than for lead teachers: while nearly half (45%) had over 10 years of maths teaching experience, around one in six lead teachers (17%) had less than five years. Like lead teachers, the vast majority reported that they were specialist maths teachers (95%). This is notable considering that nearly half of secondary schools in England rely on non-subject specialists to teach some maths lessons (Schools Week, 2022). Possible explanations are that participating schools had a higher than typical proportion of specialist maths teachers, that specialist maths teachers were deployed in Year 7, and/or that specialist maths teachers were more likely to respond to our survey. Moreover, it is possible that schools with a high proportion of specialist maths teachers were more willing to sign up to the pilot than those with fewer maths specialists.

¹⁹ The average secondary school in England in 2023 was 1054 (Homes England, 2023). We classified all schools below this average as 'small' and schools above this average as 'large'.

Table 8: Achieved sample pre-delivery survey

	Lead teachers		Year 7 teachers	
Age	N	%	N	%
20–30	13	28	17	21
31–40	15	32	26	32
41–50	10	21	26	32
Over 50	9	19	13	16
Maths teaching experience				
Less than 3 years	3	6	11	13
3–4 years	5	11	7	9
5–10 years	18	38	21	26
More than 10 years	21	45	43	52
Qualification				
Bachelor's degree or other	41	87	67	82
Master's degree	6	13	15	18
Specialist maths teacher				
Yes	43	91	78	95
No	4	9	4	5
Total	47	100	82	100

Note: Percentages may not add to 100% due to rounding.

Post-delivery survey

Lead teachers

The response rate for lead teachers was lower than for the pre-delivery survey at 62% (34 out of 55). The lower post-delivery response rate may reflect the fact that lead teachers were also invited to complete their pre-survey on the day of their training in addition to the invitation email sent by Oxford MeasurEd. In contrast, post-delivery surveys relied solely on email prompts. It may also be linked to the larger number of questions in the post-delivery survey as well as the timing; the initial invitation to complete the post-delivery survey was sent in late April, when teachers may have been occupied with GCSE exam preparation. Nonetheless, at post-delivery, the age distribution, qualification, and experience profile of lead teachers remained broadly similar to pre-delivery.

Year 7 teachers

The response rate for Year 7 teachers was lower again, at 36% (51 out of 142). The drop in response rate at post-delivery may reflect similar factors to those affecting lead teachers, but it may have been more pronounced for Year 7 teachers because they had less direct responsibility for, or ownership of, the programme. At post-delivery, the age distribution of Year 7 teachers remained relatively even, although there was a higher representation of the youngest and oldest teachers compared to the pre-delivery sample. Experience levels were somewhat more skewed towards senior staff, with 55% reporting more than ten years compared to 52% of the pre-delivery sample. The Year 7 teacher sample showed greater diversity in age and teaching experience at both timepoints, though the profile of Year 7 teachers responding to the post-delivery survey also shifted somewhat towards more experienced staff.

Table 9: Achieved sample post-delivery survey²⁰

Age	Lead teachers		Year 7 teachers	
	N	%	N	%
20–30	10	29	12	24
31–40	12	35	13	25
41–50	7	21	15	29
Over 50	5	15	11	22
Maths teaching experience				
Less than 5 years	5	15	7	14
5–10 years	13	38	16	31
More than 10 years	16	47	28	55
Qualification				
Bachelor's degree or other	31	91	44	86
Master's degree	3	9	7	14
Total	34	100	51	100

Case study schools

Case study schools were selected using primary and secondary sampling criteria to ensure the contextual diversity of pilot schools. Primary criteria included region and the proportion of FSM pupils. We had initially aimed to select an equal number of schools from the two participating regions (Yorkshire and Humber, and the North West) but selected the only two pilot schools from Yorkshire and Humber, as all remaining pilot schools were based in the North West. For FSM eligibility, we selected five schools with an above-average proportion of FSM pupils.²¹

²⁰ Subject specialism data from the post-delivery survey are suppressed to reduce the risk of disclosure.

²¹ The sampling took place in 2024 and so was based on 2023/24 figures; the percentage of pupils known to be eligible for free school meals in the North West at that time was 27.7% (DfE, 2025).

We did not stipulate quotas for the secondary sampling criteria but monitored these to maintain a diversity of pilot schools in our sample. Secondary sampling criteria included ensuring that at least two schools that had received additional support from BRS, selecting a mix of urban and rural schools proportional to the overall sample, and including diversity in classroom grouping approaches (mixed ability, prior maths attainment, prior overall attainment, or a combination).

The achieved sample comprised ten case study schools. The characteristics of the sample schools based on the sampling criteria are in Table 10.

Table 10: Achieved sample case study schools

	<i>n</i>	%
Area		
Rural	6	60
Urban	4	40
FSM status		
Above-average proportion of FSM pupils	5	50
Below-average proportion of FSM pupils	5	50
Class grouping approach		
Mixed-ability classes	3	30
Set by prior maths attainment	2	20
Set by overall attainment	1	10
Combination	3	30
Not collected	1	10
Total	10	100

Lead teachers, Year 7 teachers, and pupils were drawn from these schools for interviews, lesson observations, focus groups, and paired pupil interviews. We did not collect background characteristics of the teachers during interviews beyond years of teaching experience. We asked pupils to note down their gender on a consent form so that there was a mix of pupils identifying as male and female in our sample; overall, nine pupils identified as male and 11 as female. Beyond this, we cannot comment on the full diversity of the teacher and pupil samples. However, teaching experience among interviewed practitioners varied, and observations suggested there was a broad mix of gender and ethnicity among teachers, and an ethnically diverse pupil sample.

Feasibility of implementation

The following section explores questions around the feasibility of implementation, covering research questions (RQs) 1 to 5.

RQ1: Are intervention activities (training, ongoing support, and delivery) implemented as intended?

Key findings

The in-person training was delivered as planned. While cascading took place in most schools, it was not consistently delivered as intended. The delivery of the regional online check-in meetings deviated slightly from the delivery team's original plan, as it did not facilitate active discussions between schools. While implementation support visits were largely delivered with fidelity, this was not the case for one to one meetings, which the delivery team did not deem necessary in the intended format. Classroom sessions were delivered regularly as intended, although lead teachers and Year 7 teachers did not consistently adhere to the core components.

Training

Attendance data shows that all lead teachers received initial training, with nearly all attending the in-person sessions. Nine of 56 (16%) lead teachers could not attend any in-person sessions and instead received catch-up training from the delivery team; this was delivered in person at one school to three lead teachers and online to six (Appendix 11 Table 1). Interviews with lead teachers and the delivery team cited cover costs and sickness as key reasons some lead teachers were unable to attend the in-person training.

Observations from the in-person training and interviews with the delivery team indicated that the in-person training was delivered as planned, despite differences in group size. Unsurprisingly, perhaps given the varying group sizes (training groups ranged from 2 to 39 due to the concentration of schools in different geographies), there were minor adaptations to session timings, as there were fewer discussions with smaller groups.

The size of the training group appeared to affect the dynamics of the training. During an interview, the delivery team reflected that they had underestimated the importance of the size of the training group, which they had originally planned to be equally balanced; they highlighted that smaller groups made it more difficult to have detailed conversations and did not allow participants to have 'anonymous discussions', which were needed to make the training less reliant on the delivery team.

There's something about getting the discussion going, but there's also something about the anonymity of those discussions... I could literally hear everything they said because we were in a really small room together... There was a bit too much reliance on... me leading things... because it was such a small group. (Delivery team member—interview)

Cascading

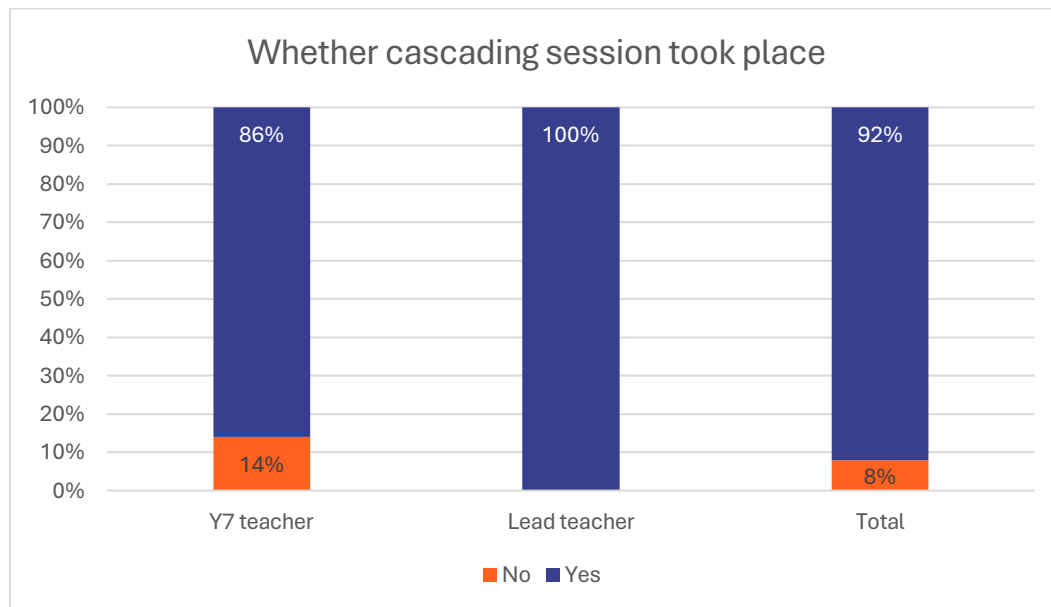
While cascading took place in most schools, it was not consistently delivered as intended. In the post-delivery survey, 78 out of 85 (92%) of Year 7 teachers and lead teachers reported that cascading occurred (lead teachers) or that they attended an internal training session (Year 7 teachers).²² The remaining 7 out of 85 respondents (8%)—all Year 7 teachers—stated they had not attended an internal training session (Figure 4, Appendix 11 Table 2). However, it was not clear from the post-delivery survey *how* they received programme information in the absence of an internal training session.

There is no evidence from the qualitative interviews with lead teachers and Year 7 teachers that cascading, or an internal training session, did *not* take place. Where Year 7 teachers joined the school after the internal training session had taken

²² We asked Year 7 teachers about 'internal training' rather than using the word 'cascading', as we assumed they would not all be familiar with the meaning of the word 'cascading' in this context.

place, they reported receiving separate sessions and/or one-to-one sessions with the lead teacher to learn about the programme.

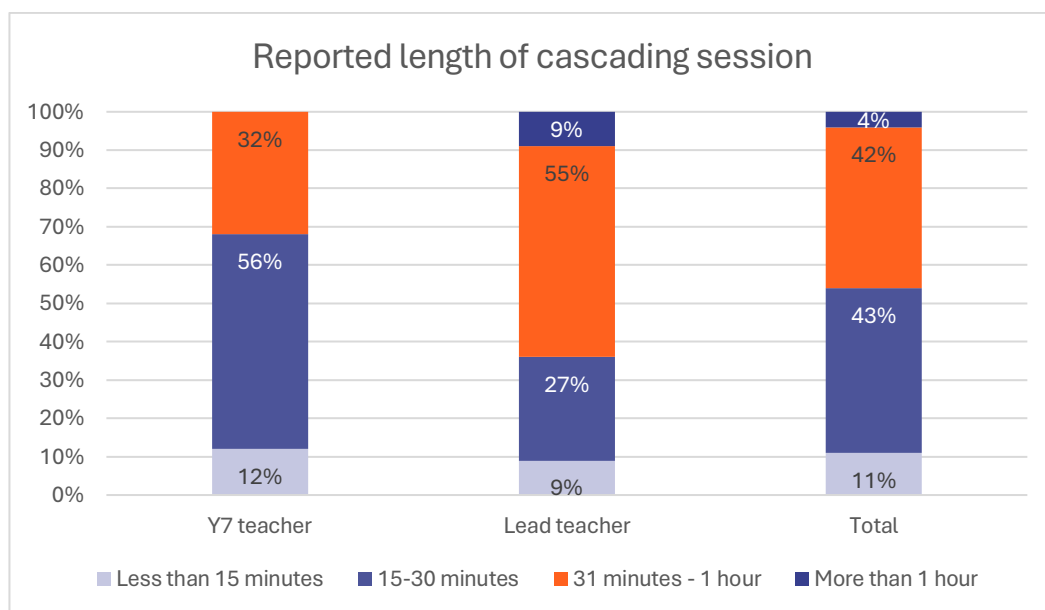
Figure 4: Reported occurrence of cascading session by role on the pilot



Base: All respondents of the post-delivery survey (Total, $N = 85$, Y7 teachers, $n = 51$, lead teachers, $n = 34$)

Source: Post-delivery survey

Figure 5: Reported cascading session length by role on the pilot



Base: All respondents of the post-delivery survey who reported and recalled cascading session length (Total, $N = 74$, Y7 teachers, $n = 41$, lead teachers, $n = 33$)

Source: Post-delivery survey

Where cascading did occur, session duration varied. The intended session length was at least one hour; however, 23 out of 41 (56%) Year 7 teachers and 9 out of 33 (27%) lead teachers responding to the post-delivery survey reported sessions lasting only 15–30 minutes, 13 out of 41 (32%) Year 7 teachers and 18 out of 33 (55%) lead teachers reported sessions being between 31–60 minutes, and only 5 out of 41 (12%) Year 7 teachers and 3 out of 33 (9%) lead teachers reported sessions being less than 15 minutes. No Year 7 teachers and the remaining 3 out of 33 (9%) lead teachers reported that sessions lasted more than one hour (Figure 5, Appendix 11 Table 3). Qualitative interviews with lead teachers similarly revealed variation in the length of cascading sessions, with one session lasting up to two hours due to lead teachers asking their staff to engage in role-play, which required more time. In the survey, lead teachers were significantly more likely than Year 7 teachers to describe longer sessions ($p = 0.021$); it is possible that this variation is driven by social desirability bias, since lead teachers were familiar with the requirements of the length of cascading sessions. Another explanation is that they were more likely to correctly remember the duration of their cascading sessions, as it was a key part of their role. However, given the small sample size, this finding must be interpreted cautiously.

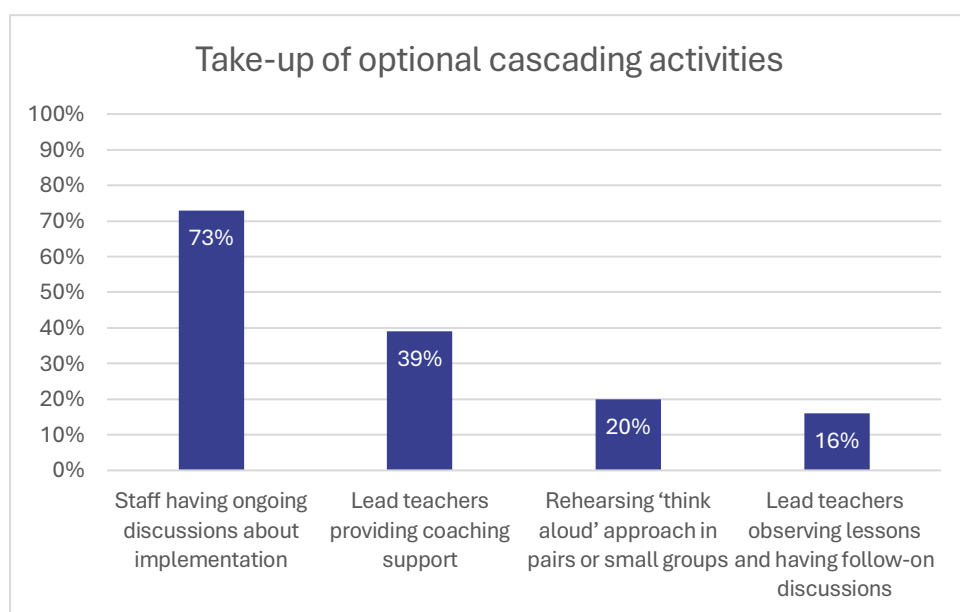
There was also variation across schools regarding their engagement with mandatory cascading activities. This included lead teachers delivering a one-hour presentation to Year 7 teachers before delivery; the use of videos and resources as part of the session; and ringfencing an hour for Year 7 teachers to engage with the resources before delivering the sessions. Among lead teachers and Year 7 teachers responding to the post-delivery survey who reported a cascading session had taken place, 68 out of 78 (87%) reported a presentation from lead teachers; however, only 38 out of 78 (49%) reported watching the video from the delivery team (BRS), and just 20 out of 78 (26%) had received the handbook (Appendix 11 Table 4). Given the limited survey response rates and the risk of a biased sample, these findings need to be treated with caution.

Interviews with lead teachers and focus groups with Year 7 teachers highlighted that videos were either watched during the cascading sessions or shared with staff to watch individually, for instance, where there was limited time available for the session. Finally, only 30 out of 84 (36%) lead teachers and Year 7 teachers reported being given an hour to explore lesson resources in advance. Also, 40 out of 84 (48%) said they did not have this time. The remaining 14 out of 84 (16%) were unsure whether this time had been set aside (Appendix 11 Table 5). The extent to which teachers reported having one hour set aside to engage with resources before delivery varied by school size, with Year 7 teachers and lead teachers from larger schools more likely to report that an hour was not set aside to explore resources than Year 7 teachers and lead teachers from smaller schools ($p = 0.010$, Appendix 11 Table 5). A possible explanation for this difference might be that

there was more lead teacher ‘oversight’ of the programme in small schools where fewer teachers were delivering the programme. However, given the small sample size, this finding must be interpreted cautiously.

There was variation among schools regarding the take-up of optional cascading activities. In the post-delivery survey, a few Year 7 teachers and lead teachers reported engaging in recommended activities designed to support fidelity within schools. Only 33 out of 85 (39%) reported lead teachers providing coaching support, 17 out of 85 (20%) reported rehearsing the ‘think aloud’ approach in pairs or small groups, and 14 out of 85 (16%) reported lead teachers observing lessons with follow-up discussions (Figure 6, Appendix 11 Table 6). Some lead teachers noted in interviews that they modelled sessions in a class or gave Year 7 teachers the opportunity to practice elements of the sessions themselves. Where lead teachers observed TM sessions, this took different forms. Examples included lead teachers observing each Year 7 teacher at least once or doing occasional drop-ins. Lead teachers who explained in interviews that they did not observe Year 7 teachers delivering classroom sessions cited logistical barriers, such as lead teachers teaching at the same time as Year 7 teachers.

Figure 6: Take-up of optional cascading activities



Base: All respondents of the post-delivery survey (Total, $N = 85$, Y7 teachers, $n = 51$, lead teachers, $n = 34$)

Source: Post-delivery survey

The most widespread activity to monitor fidelity was ongoing departmental discussions about implementation, reported by 62 out of 85 (73%) respondents to the post-delivery survey (Figure 6, Appendix 11 Table 6). This was reflected in interviews and focus groups with lead teachers and Year 7 teachers. Some schools chose to discuss ongoing implementation during regular team meetings. More commonly, though, Year 7 teachers discussed delivery with colleagues and their lead teachers in an ‘ad hoc’ way, such as during informal catch-ups in the staff room. There was a view among lead teachers and Year 7 teachers that packed meeting agendas could make it difficult for TM to be a standing item on the meeting agenda.

Ongoing support

Attendance at the two regional online check-ins was mixed, although nearly all schools had at least one representative at each session. Attendance data collected by the delivery team showed that 27 out of 56 lead teachers (48%) attended the first regional online check-in, and 29 out of 56 (52%) attended the second (Appendix 11 Table 1). While both lead teachers from each school were expected to participate, the delivery team explained in an interview that it was sufficient for one representative per school to be present. Even so, at the first regional check-in, three schools (10%) had no representatives, while four (14%) had both lead teachers present. At the second, two schools (7%) had no representatives and only one (4%) had both.

Qualitative interviews with lead teachers and the delivery team identified practical reasons for non-attendance, including lead teachers needing to attend parents' evenings or timetabling clashes. A lead teacher at one school recalled having been unaware of the check-in meeting, as they had not received an email about the event and so did not attend.

I did not get the emails initially... the Head of Department was getting all those emails and then he was forwarding them on to me, but I didn't get any notification about an online check-in meeting. (Lead teacher—interview)

During an interview, the delivery team explained that they sent a recording of the session to all non-attendees and required them to fill in a form once the video was watched or confirm this via email.

Observations showed that the delivery of the regional online check-in meetings deviated slightly from the delivery team's original plan, as it did not facilitate active discussions between schools. A delivery team member explained that they had decided against having a more interactive session because they wanted to avoid an *'unfocused... talking shop'*, although lead teachers were given the option of asking questions using the chat function. The delivery team were clear during interviews that they wanted to ensure the sessions communicated key messages about delivery (for example, the core components) as well as sharing lead teachers' experiences.

We're a number of weeks in now and we just felt that with Christmas approaching... the chance of an implementation dip is quite high. So we wanted to...refocus those key messages. We also wanted to share with the people on the call the initial findings from the delivery logs...We wanted to make sure that people felt that we were using those for something. (Delivery team member—interview)

While a key part of the sessions included the delivery team revisiting the key components, they presented some feedback from lead teachers from the delivery logs in the sessions. Moreover, they gauged lead teachers' experiences of delivery through the use of Mentimeter,²³ focusing primarily on their views of the problems. This included questions such as what problems generated the most discussion or what problems pupils found most accessible.

Implementation support visits were largely delivered with fidelity; the delivery team reported in interviews that they were delivered to six schools who had reported challenges with implementation via the delivery log. While the study plan does not specify what proportion of schools receive an implementation visit from BRS, the delivery team explained during the learning workshop that they aimed to visit around 20% of all pilot schools, which is consistent with the number of schools visited. Visits lasted between two and four hours (Appendix 11 Table 7), in line with the expected duration of visits outlined in the study plan, which specifies that the duration of visits is tailored to each school's needs but typically around half a day long.

The delivery team confirmed in follow-up correspondence via email in December 2025, that they adhered to the two criteria which triggered a visit; namely, 1) schools failing to complete two delivery logs in a row; and 2) after week 3, the classroom sessions taking longer than stipulated as per the delivery log. However, they explained that they only selected those settings where sessions were routinely reported as taking longer than 20 minutes, as none of the settings failed to complete two delivery logs in a row during the first eight weeks of implementation.

A key part of the implementation support visit includes the delivery team observing sessions and providing feedback. All lead teachers who reported receiving an implementation support visit in the post-delivery survey reported the delivery team observing classroom sessions, while 5 out of 7 (71%) reported feedback being given by the delivery team following the observation during the visit (Appendix 11 Table 7).

The delivery team explained in interviews that for each visit, they used a form to record notes on delivery, followed by discussions with the lead teacher. With the exception of one school, the delivery team reported observing at least two sessions; this enabled them to capture variability and consistency between teachers and provide more meaningful feedback and suggestions to the lead teacher.

²³ Mentimeter is an online polling software which can be used to share live or asynchronous polls, quizzes, and word clouds.

The online one-to-one meeting with BRS did not take place as intended, with only 4 out of 34 (12%) of the lead teachers reporting having had such a meeting in the post-delivery survey (Appendix 11 Table 1). A delivery team member acknowledged in an interview that they only held short meetings with the same six schools who also received the implementation support visit following concerns arising from the logs. They did not deem it necessary to have individual 30-minute meetings with every school, specifically those that adhered to the prescribed session duration.

Those schools that were saying it was consistently taking around 20 minutes to deliver, it was happening in all classrooms, the teachers were reporting that it was successful. It was just unnecessary to engage with those schools at a deeper level. (Delivery team member—interview)

Classroom sessions

While classroom sessions were delivered regularly, implementation was uneven across schools, lead teachers, and Year 7 teachers. In the post-delivery survey, 73 out of 85 (86%) Year 7 teachers and lead teachers reported that sessions were delivered every week (Appendix 11 Table 8). However, the delivery logs submitted by lead teachers and shared with the delivery team provide a contrasting picture, suggesting lower uptake than teacher self-reports. Across all schools over the course of 20 weeks, lead teachers reported all Year 7 teachers in their schools had used the programme materials in the last two weeks in an average of only 147 out of 267 (55%) reporting instances (Appendix 11 Table 9).²⁴

Lead teachers and Year 7 teachers did not consistently adhere to the prescribed classroom session length. Post-delivery survey data shows that only 13 out of 85 (15%) Year 7 teachers and lead teachers reported that all sessions lasted 20 minutes (Appendix 11 Table 10). Delivery logs corroborate this variation: in 149 out of 261 (57%) reporting instances, sessions were reported as lasting up to 20 minutes, in 103 out of 261 (39%) of reporting instances, they were reported as lasting 21–30 minutes, and in only 9 out of 261 (3%) of reporting instances, sessions were reported as lasting longer than 30 minutes (Appendix 11 Table 11). Lead teachers were significantly more likely than Year 7 teachers to report longer sessions ($p = 0.004$, Appendix 11 Table 12); it is possible that this difference was driven by lead teachers being more aware of and implementing all core components than Year 7 teachers. It could also be driven by social desirability. However, given the small sample size, this finding must be interpreted cautiously.

Lead teachers and Year 7 teachers explained in interviews and focus groups that the perceived complexity of the problem or the use of a visualiser for the debrief—a piece of technology similar to an overhead projector and used for showing work to the whole class—meant that sessions could take longer than 20 minutes.

I would say it takes about 20–25 minutes for me to do the whole thing properly, maybe even a bit longer if I have chosen somebody's work and put it under the visualiser, and we have talked about it. (Year 7 teacher—focus group)

Lead teachers and Year 7 teachers did not consistently adhere to all core components when delivering classroom sessions. In the post-delivery survey, 81 out of 85 (95%) respondents reported that problems were used in the prescribed order (Appendix 11 Table 13), and 76 out of 85 (89%) reported that independent problem-solving took place in every session. A smaller majority of 67 out of 85 (79%) reported that pupils discussed and compared solutions and approaches during every session (debrief), while 65 out of 85 (76%) said that they always modelled their own thinking using a 'think aloud' (Appendix 11 Table 14). Observations highlighted that teachers used the 'think aloud' approach less towards the latter stages of the programme, where pupils were more accustomed to the process of approaching a problem. However, the debrief took place in each of the classroom sessions we observed during our visits and either included teachers calling on individual students to share their thinking with the whole class (for example, via a visualiser) or pupils sharing their thinking with their neighbour followed by a whole-class discussion. Only 54 out of 85 (64%) reported consistent use of prompts by pupils to guide their thinking (Appendix 11 Table 14).

²⁴ The delivery team explained that the discrepancy between the survey data and delivery logs could be attributed partly to a small number of schools discontinuing the programme for their lowest-attaining ('bottom set') pupils. This meant lead teachers could not report full weekly use.

Interviews and focus groups with lead teachers and Year 7 teachers revealed that pupils had already gotten used to prompts. This meant that there was no perceived need to refer to these explicitly to tackle a mathematical problem, suggesting that the process of self-questioning was becoming established among pupils.

They did more so at the beginning [use the prompts], but they have now gotten more into a routine that they know what they are, and they know how to tackle the question a bit better than at the beginning. (Year 7 teacher—focus group)

Teachers reported modifying the questions to suit their pupils' needs, thereby adapting delivery. This included providing extension tasks for 'high-ability' pupils to maintain their engagement if they finished sooner, for instance, by asking them to come up with their own problem-solving questions. During interviews, Year 7 teachers also reported providing more scaffolding or rewording questions for lower-ability pupils to ensure they could start the process of solving the problem.

RQ2: Are intervention components of an appropriate quality?

Key findings

Lead teachers reported high levels in both understanding and confidence following initial training from BRS and praised the accessibility of the training delivery. The regional check-in meetings were useful for providing an opportunity to learn about other schools' challenges but were not seen as sufficiently interactive. There was a view that some problems were either too easy for high attainers or too difficult for pupils less secure in maths.

Initial training

Lead teachers reported high levels in both understanding and confidence following initial training from BRS. Altogether, over 90% of lead teachers responding to the post-delivery survey said they understood either 'quite a bit' or 'a lot' about how to deliver the TM programme. A similarly high proportion (over 90%) reported feeling 'confident' or 'very confident' in their ability to deliver the programme. Very few reported only a limited understanding or low confidence in delivering the programme (Appendix 11 Table 15).²⁵

During interviews, lead teachers cited the accessibility of the delivery from BRS, the emphasis on being able to rehearse key strategies such as the 'think aloud', and the clear rationale for why the programme was needed as key reasons that they felt confident about delivery following the initial training. Lead teachers praised the clarity of guidance they received, which helped them feel prepared.

I was pleasantly surprised... that's by far the most productive one I've ever been to, where they actually had a plan, and they delivered what they wanted to and told us exactly what they expect from us, which was good. (Lead teacher—interview)

We came not knowing anything about it, but we left with a lot of information and a lot more confidence in what we were doing. (Lead teacher—interview)

Lead teachers praised the time put aside to engage with the resources, specifically the question bank of weekly problems. This helped bring the problems alive and gave lead teachers a better understanding of how they would practically work in a classroom environment.

[O]ne real positive of those training days...was that opportunity for people to actually have a go at the problem, talk about the problems, talk about this is what I think the pupils would do if they tried this problem. Because what we were trying to do there is model the conversations that we would like to happen in schools. (Lead teacher—interview)

²⁵ Exact figures have been suppressed to reduce the risk of disclosure.

However, there was a view among some lead teachers that more time could have been set aside to focus on the debrief and the prompt questions. This was considered important because the debrief and the prompt questions set the programme apart from a typical maths lesson.

That [debrief and prompts] is the part that is different to what you would typically do in a maths lesson. We don't really do much group work, and we don't really do much problem-solving at all in Key Stage 3. (Lead teacher—interview)

Resources

Year 7 teachers' views of available resources for teaching problem-solving improved substantially after they had accessed the programme materials for TM, which suggests that the materials were useful. Before delivery, only 5 out of 85 (6%) of all lead teachers and Year 7 teachers responding to the pre-delivery survey considered existing problem-solving resources 'very useful', and 33 out of 85 (39%) described them as 'only slightly useful' or 'not useful'. After the end of programme delivery, 24 out of 85 (28%) lead teachers and Year 7 teachers responding to the post-delivery survey rated resources as 'very useful', and the proportion finding them 'only slightly useful' or 'not useful' dropped to 11 out of 85 (13%) (Appendix 11 Table 16). However, the findings need to be treated with caution. It should be noted that the improvement in perceived usefulness of resources may not be solely attributable to the TM programme materials, as schools and teachers may also have acquired or made greater use of other problem-solving resources during the same period. Further, it is possible that those participants who found the resources less useful did not respond to the post-delivery survey, particularly given the lower response rate in the post-delivery survey compared to the pre-delivery survey.

Evidence from the delivery logs supports the survey findings on the usefulness of resources. In the delivery logs, lead teachers reported in 174 out of 266 (65%) reporting instances that 'all' Year 7 teachers found the materials useful and beneficial, while a further 89 out of 266 (33%) of the time the lead teachers said that 'most' did (Appendix 11 Table 17). This pattern was consistent across weeks of delivery, with little week-to-week variation.

Interviews with lead teachers and focus groups with Year 7 teachers suggested that resources, notably the slides, the teacher notes, and the website where the materials were stored, were easy to follow and to navigate (see RQ4). Moreover, the non-technical language used meant that the resources were also accessible for non-subject specialists.

While some lead teachers praised the selected problems for encouraging pupils to think 'hard' and 'deep', other lead teachers and Year 7 teachers explained that the problems could be either too easy for high attainers or too difficult for pupils less secure in maths, leading to some disengagement among pupils (see RQ9). This concern tended to be articulated by lead teachers and Year 7 teachers in schools where there was a high proportion of pupils who struggled with maths as well as pupils with special educational needs and disabilities (SEND) and pupils with English as an additional language (EAL). While pupils who were less secure in maths generally found problem-solving difficult, a view from teachers was that some pupils with SEND particularly struggled because of the lack of visuals. Relatedly, some problems were perceived to be overly 'wordy', which presented a challenge for EAL pupils.

The delivery team noted in an email correspondence following programme delivery that a small number of schools discontinued the programme for their lowest attaining ('bottom set') pupils, further suggesting that some schools found the problems less suitable for pupils who struggled in maths.

Ongoing support

Lead teachers' views on the usefulness of ongoing support were mixed. In the post-delivery survey, 25 out of 32 lead teachers (78%) rated all of the training and support received as 'useful' or 'very useful', but a notable minority—7 out of 32 (22%)—reported finding it 'only slightly useful' or 'not useful' (Appendix 11 Table 18).

Some lead teachers explained in interviews that they enjoyed the opportunity to learn from other schools during the regional check-in meetings, as this provided insights into whether other schools had similar challenges. Other lead teachers were less positive about the check-in meetings and explained that a different format would have been more purposeful, such as inviting questions from participants directly. However, as a delivery team member pointed out during an interview, the

delivery team deliberately decided to make the session less interactive in order to maintain a clear focus on the session aims (see RQ1).

Qualitative interviews collected limited data on the usefulness of the one-to-one meetings and implementation support visits, as these happened at only a few schools. However, the delivery team were positive about these sessions, as they were able to share concrete suggestions with the lead teacher. At the same time, they questioned how ‘representative’ the sessions they observed were to help them gauge a comprehensive understanding of delivery, particularly at schools where multiple Year 7 teachers delivered the sessions. Time constraints meant that the delivery team were limited in the number of sessions they could observe at each school they visited as part of the implementation support visit.

Dosage

While there was no indication that the programme was too long, some lead teachers and Year 7 teachers wanted the programme to run for longer than 20 weeks, largely to sustain practice and to continue a routine. For instance, a Year 7 teacher highlighted the importance of establishing a routine in order for pupils to know how to approach a mathematical problem when they were in Key Stage 4. Others believed 20 weeks was sufficient due to the perceived repetitive nature of some of the problems; for instance, a Year 7 teacher explained that pupil engagement had declined towards the end of the programme but acknowledged that more varied types of problems could help extend the programme for more than 20 weeks while maintaining pupil engagement.

As with the duration of the programme, there was no consensus on the **lead teachers’ and Year 7 teachers’ views on the appropriate session duration**. A view was that 20 minutes per week was suitable, with a lead teacher noting that any less time would not be enough to help pupils improve their resilience. Some Year 7 teachers wanted more flexibility regarding the timings because they believed the duration of sessions was dependent on pupil ability. For instance, a Year 7 teacher noted that 20 minutes was not enough for their bottom set and that a flexible approach was needed to engage all pupils.

It has to be a flexible approach because if the kids are going to get something out of it, then you can't say, alright it's 20 minutes and you didn't solve that problem, so next week we're going to do a different one. That would disengage them so badly. You can't run a problem-solving session and not solve the problem because otherwise that just proves to the children that they can't problem solve, and that would make them give up much faster. (Year 7 teacher—focus group)

The delivery team were clear that sessions should not exceed 20 minutes. During the learning workshop, they explained that the duration was appropriate because they saw the weekly sessions as part of classroom teaching rather than an intervention. And a longer session risked schools being unable to find time to deliver it.

The rationale for the 20 minutes was [that] if you make it significantly longer than that, schools won't be able to find the time to deliver it...because it isn't an intervention. It is part of classroom teaching. You've got to be really careful not to take too much time away from teachers or they will just say, 'well, I can't do that'. So, I think the 20 minutes...has to stay. (Delivery team member—learning workshop)

RQ3: Are the resources, training, and support provided appropriate for schools with above-average numbers of FSM eligible pupils?

Key findings

Survey and qualitative findings provide no indication that training, resources, and materials were experienced differently by schools with above-average numbers of FSM eligible pupils.

There is no evidence that the resources, training, and support provided are less suitable for schools with above-average number of FSM-eligible pupils. There were no significant differences between higher- and lower-FSM schools in how Year 7 teachers and lead teachers viewed the training, resources, or their own confidence and understanding, and both groups showed similar positive improvements following delivery (Appendix 11 Table 19–21).

Qualitative findings provide no indication that training, resources, and materials were experienced differently by schools with above-average numbers of FSM eligible pupils. While there were concerns around the suitability of the problems for pupils less secure in maths (see RQ2), this was a concern that was echoed in schools with a higher and lower proportion of FSM pupils.

RQ4: What are the barriers and facilitators to successful implementation and take-up of the programme, including in schools with above-average numbers of FSM eligible pupils?

Key findings

Lead teachers praised the trainers' expertise and ability to answer questions, making the training and programme accessible. A lack of time impacted lead teachers' ability to cascade the training in their school, with some schools unable to dedicate the required time to cascade the training. Session slides and scripts helped classroom delivery, ensuring consistency in how sessions were delivered throughout the programme and between teachers.

The following section discusses the barriers and facilitators to successful implementation of TM, including training, cascading, ongoing support, and classroom delivery.

Training

Resources introduced during the training day – slides, the training handbook, and the website – helped lead teachers access and take part in training. In the post-delivery survey, 25 out of 34 (74%) lead teachers highlighted the usefulness of resources in supporting their access to training (Appendix 11 Table 22), which was echoed in interviews with lead teachers; they felt positive about the resources, which were accessible and readily available. As a result, they increased the lead teachers' confidence about how to deliver the sessions.

We left... with a lot of information and a lot more confidence... we knew exactly what we were doing and how we were going to deliver. (Lead teacher—interview)

Support from the delivery team helped lead teachers to access and understand the training. In the post-delivery survey, 19 out of 34 (56%) lead teachers responding pointed to support from BRS as a facilitator to training (Appendix 11 Table 22). This was reflected in interviews with lead teachers, who praised the clarity of delivery and the trainers' expertise and ability to answer questions.

Moreover, 9 out of 34 (26%) lead teachers responding to the post-delivery survey cited SLT support as a facilitator to accessing training (Appendix 11 Table 22). Although respondents did not elaborate on this in the survey, it is possible that they referred to members of SLT supporting the programme and granting them permission to attend an in-person training session.

Different group sizes affected delivery. While the delivery team explained in interviews that training groups worked well together, they noted how differences in group sizes impacted delivery (see RQ1). For instance, one of the training sessions we observed only included three schools with one participant each; this resulted in more limited opportunities for schools to share and learn from each other and a stronger reliance on the delivery team to lead the session (see RQ1).

Cascading

The most frequently cited facilitator to delivering cascading by lead teachers in the post-delivery survey was the initial BRS training, with 23 out of 34 (68%) lead teachers identifying this as a facilitator (Appendix 11 Table 23), reflecting the finding from the post-delivery survey that lead teachers felt sufficiently prepared to deliver the training internally (see RQ2).

The timing of the sessions was seen as another key factor in ensuring that cascading took place, with favourable timing of the training session being cited by 37 out of 44 (84%) Year 7 teachers in the same survey (Appendix 11 Table 23). This

suggests that lead teachers being provided with time at the end of their in-person training session (see below) to schedule a cascading session ensured that cascading could take place in their school.

During interviews, lead teachers explained **that the cascading resources (the cascading video and PowerPoint) were comprehensive and made preparation more straightforward for them**, even where time was more limited. Although fewer respondents in the post-delivery survey identified resources such as the handbook (18 out of 78, 23%) or video (29 out of 78, 37%) as facilitators, they appeared to resonate most with lead teachers, who made up 72% of those selecting each resource as a facilitator (Appendix 11 Table 23).²⁶ It is very likely that this is due to lead teachers actively drawing on the resources to deliver the training to Year 7 teachers, compared to Year 7 teachers who received the resources as part of cascading.

Time impacted lead teachers' ability to cascade the training in their school. In the post-delivery survey, 11 out of 34 (32%) lead teachers reported lack of time as a barrier (Appendix 11 Table 24). Interviews with lead teachers further indicated that some schools were unable to dedicate an hour to cascade the training as recommended by the programme, for instance, due to infrequent departmental meetings, limited staff availability, and other CPD commitments. This was despite the delivery team asking lead teachers to schedule the cascading session during the in-person training session:

[W]e had time during the training day for the delivery leads to get the calendar open, have a look at when that cascading session was going to take place... The fact that we allowed them that time just highlighted the importance of that. (Delivery team member—interview)

Interviews with lead teachers suggest that SLTs did not consistently provide lead teachers with the time to cascade.

In one example, a lead teacher reflected that even though they had asked their head of maths to put the whole team meeting aside for the cascading session, they had only been given ten minutes for the session itself. However, there were other examples of lead teachers highlighting that despite difficulties of arranging the sessions, SLTs ensured they could fulfil the cascading requirements, for instance by scheduling two shorter cascading sessions rather than one long one.

Ongoing support

Interviews with lead teachers revealed that while the online mode of the regional check-in meetings made participation accessible, the large number of participants hindered engagement between participants; for instance, lead teachers highlighted that the lack of opportunity to share and discuss challenges among different schools implementing the programme meant the meeting was less helpful than it could have been (see RQ2). Qualitative data on barriers and facilitators to ongoing support was limited to these regional check-in meetings.

Classroom delivery

Most lead teachers and Year 7 teachers felt very positive about the resources, including the weekly session slides, the teacher notes including the script, worked examples, prompts, and the weekly printable problem. In the post-delivery survey, 74 out of 85 (87%) of all respondents reported resources as a facilitator to classroom delivery (Figure 7). Interviews with lead teachers and focus groups with Year 7 teachers revealed that they appreciated the availability of a package of multiple resources that was ready for them to pick up and use (Appendix 11 Table 25). This meant little extra planning was needed before delivering the sessions, which was noted to be particularly helpful for non-specialist teachers.

Interviews and focus groups highlighted that lead teachers and teachers perceived the session slides to be clear, and the integration of the prompt questions facilitated discussion amongst pupils in the classroom. Teachers and lead teachers particularly liked that the teacher notes and script helped ensure some consistency in how the sessions were delivered.

²⁶ Differences in facilitators reported by lead teachers and Year 7 teachers may be due to the way in which the question was posed to the two groups. Lead teachers were asked about facilitators and barriers to delivering cascading, while Year 7 teachers were asked about facilitators and barriers to accessing the internal training session.

I would like reading the script to help me to think about the problem. And it's also really helpful that the solutions are there as well because it is time consuming if you had to do it and potentially could make a silly mistake as well. So that's been good. (Year 7 teacher—focus group)

All of those prompts and bullet points on the PowerPoint and in the script are really useful because you can talk through that with the students. (Year 7 teacher—focus group)

Year 7 teachers (24 out of 51, 47%) responding to the post-delivery survey valued support from lead teachers. Younger Year 7 teachers were more likely to find lead teacher support useful than older teachers, which may reflect their more limited teaching experience. Few post-delivery respondents cited SLT support (8 out of 85, 9%) or rehearsal opportunities (12 out of 85, 14%) as facilitators (Appendix 11 Table 25). It is likely that this is due to SLT being less involved in classroom delivery as well as schools providing few rehearsal opportunities during delivery. However, teachers in smaller schools (8 out of 31, 26%) were more likely than those in larger ones (4 out of 45, 9%) to view rehearsal opportunities as beneficial ($p = 0.047$, Appendix 11 Table 25), potentially because teachers at smaller schools had fewer colleagues to have ongoing discussions with and rehearsal opportunities provided an alternative way of supporting delivery. However, given the small sample size, this finding must be interpreted cautiously.

In the post-delivery survey, 37 out of 85 respondents (44%) identified ongoing discussions with colleagues about implementation as helpful (Appendix 11 Table 25, Figure 7). This was mirrored in focus groups with Year 7 teachers, who noted that ongoing informal discussion with their colleagues helped them refine how to pitch their sessions.

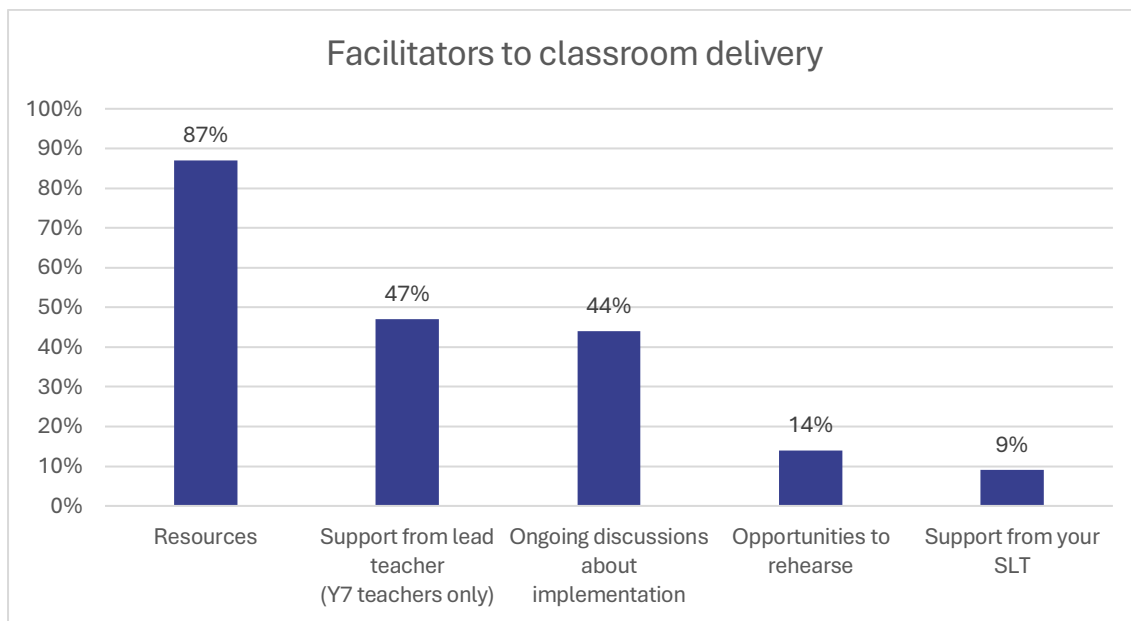
Time could be a barrier to classroom delivery. Time constraints were the most commonly reported barrier in the post-delivery survey (25 out of 85, 29%, Figure 8). Year 7 teachers (20 out of 51, 39%) were more likely than lead teachers (5 out of 34, 15%) to report lack of time as a barrier ($p = 0.015$) (Appendix 11 Table 26); this may be because lead teachers were more confident in delivering the session within the prescribed 20 minutes. However, given the small sample size, this finding must be interpreted cautiously. Focus groups with Year 7 teachers highlighted that a perceived lack of time meant that certain components could be left out, notably the debrief, as this tended to happen towards the end of sessions. During interviews and focus groups, lead teachers and Year 7 teachers noted that the problems had varying levels of difficulty which impacted the length of time spent on each session.

A couple of the problems have been a little too easy, and they were solved very quickly by some of the class. The week 16 one was very challenging for the group I take. (Lead teacher—interview)

The session varies for every single problem. I have some problems [that] we've done in less than five minutes. Really quick, even with some students who weren't even the highest of abilities. However, there's been other problems which have taken the whole lesson, but the students have still not understood it properly. So, it is such a wide variety... It would be nice if the sessions were timed more consistently... but it's not turned out like that, and I've had to be quite flexible with it. (Year 7 teacher—focus group)

Pupil ability could impact delivery. Lead teachers and Year 7 teachers expressed that a lack of differentiation in the mathematical problems meant that some questions were either too easy or too difficult (see RQ2). This challenge is echoed in the post-delivery survey, where 15 out of 85 (18%) respondents cited mixed-ability classrooms as a barrier to classroom delivery (Figure 8, Appendix 11 Table 26).

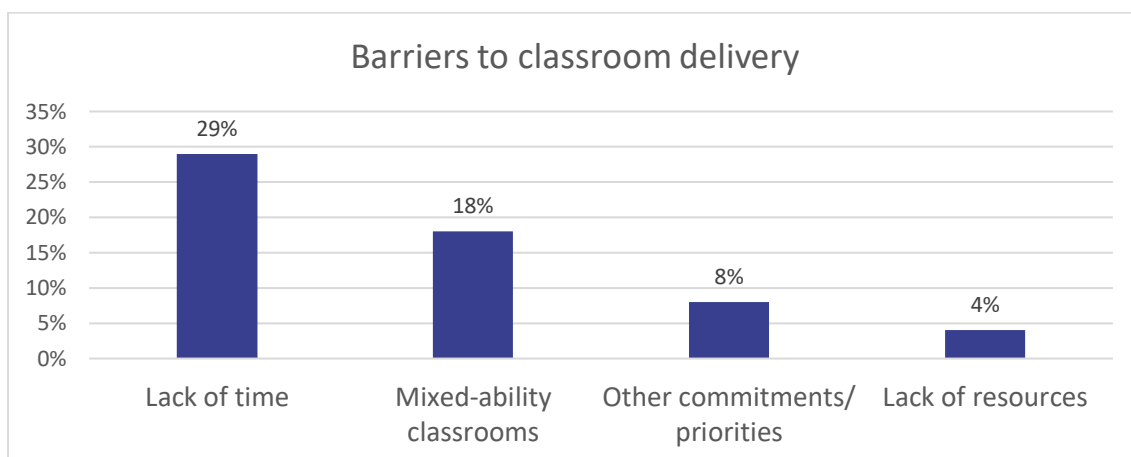
Figure 7: Reported facilitators to classroom delivery



Base: All respondents of the post-delivery survey (Total, $N = 85$, Y7 teachers, $n = 51$, lead teachers, $n = 34$)

Source: Post-delivery survey

Figure 8: Reported barriers to classroom delivery



Base: All respondents of the post-delivery survey (Total, $N = 85$, Y7 teachers, $n = 51$, lead teachers, $n = 34$)

Source: Post-delivery survey

RQ5: Are there any unintended consequences of the programme's delivery?

Key findings

Lead teachers and Year 7 teachers believed the programme resulted in a better understanding of their pupils' abilities and improved their relationships with pupils. The programme was also perceived to have improved pupils' communication skills, teamwork, and ability to problem solve beyond maths. More negative unintended consequences included lead teachers and Year 7 teachers reporting that the programme could reduce the time available for other curriculum content.

Programme delivery was perceived to have resulted in a range of positive unintended consequences at a teacher and pupil level. In the post-delivery survey, 49 out of 85 (58%) Year 7 and lead teachers reported that the programme gave them a better understanding of their pupils' abilities, and a third (33%) felt it improved their relationships with pupils (Appendix 11 Table 27). This was echoed by lead teachers, who identified that building relationships with pupils was helped by having conversations about the process of solving problems and different alternatives which were different from 'business as usual'.

The programme was perceived to have improved pupils' communication skills. In the post-delivery survey, 56 out of 85 (66%) lead teachers and Year 7 teachers reported improved communication as a benefit (Appendix 11 Table 28). In qualitative encounters with the delivery team, pupils, Year 7 teachers, and lead teachers, the debrief element of the TM sessions was particularly highlighted as improving pupil oracy. For instance, a pupil explained how their communication skills had improved by being asked to regularly present their approach to problem-solving. Similarly, a lead teacher observed a marked change in pupils' ability to express themselves clearly and concisely, which they believe had been helped by the specific prompts.

They don't waffle anymore. They're quite on target with their discussions now. (Lead teacher—interview)

Pupils' ability to work with peers was also reported to have improved through the programme. In the post-delivery survey, 31 out of 85 (36%) lead teachers and Year 7 teachers identified better teamwork as a positive unintended consequence (Appendix 11 Table 28). Lead teachers, Year 7 teachers, and pupils highlighted this in interviews and focus groups, noting that pupils became better at working with each other, for instance, by working with different people and experiencing being the stronger or weaker person in the pair. A lead teacher pointed out that peer working was particularly beneficial for pupils low on confidence, as their confidence was less likely to 'break' if corrected by a fellow pupil.

Enhanced problem-solving skills beyond maths was reported as another positive unintended consequence from programme participation by 32 out of 85 (38%) lead teachers and Year 7 teachers responding to the post-delivery survey (Appendix 11 Table 28). Year 7 teachers and pupils shared in focus groups and interviews that they felt problem-solving skills could be applied in other subject areas outside of maths. For instance, Year 7 teachers noted that pupils could benefit from techniques they had learned, such as reading the question twice and picking out the key information. Pupils noted that the ability to 'think outside the box' and consider different approaches to answering a question helped them with other subjects such as science.

Lead teachers and Year 7 teachers reported in interviews and focus groups that they started using TM strategies with other year groups, largely because they felt more confident and knowledgeable about teaching problem-solving and saw the benefits of the strategies in their own practice (see RQ8).

However, lead teachers and Year 7 teachers also reported more negative unintended consequences at a curriculum level. For instance, **some lead teachers and Year 7 teachers reported that the programme had a negative impact on the curriculum content.** In the post-delivery survey, 38 out of 85 (45%) respondents reported reduced time available for other curriculum content as a negative unintended consequence (Appendix 11 Table 28). It was apparent during focus groups with Year 7 teachers that this generated some concern.

We've found that it has affected our teaching time and about how much we've got through the same content in terms of our scheme of work, just because we're using one out of three lessons for problem-solving. (Year 7 teacher)

There was also a perception that the programme risked positioning problem-solving as a standalone aspect of the maths curriculum rather than an integrated part. In the post-delivery survey, 32 out of 85 (38%) of all lead teachers and Year 7 teachers agreed with this statement, with less experienced teachers more likely to report this as an issue ($p = 0.013$, Appendix 11 Table 27). Less experienced teachers might have perceived the weekly 20-minute slot as both difficult to allocate within an already crowded timetable and as reducing the need to integrate problem-solving into other lessons, especially if they already found teaching problem-solving challenging. However, given the small sample size, this finding must be interpreted cautiously.

Evidence of success

Table 11 includes a summary of the evidence of success presented for the success indicators related to feasibility of implementation, covering RQs 1–4.

Table 11: Success indicators related to feasibility of implementation

Success indicators	Evidence of success
Programme activities (training/ongoing support/in-class delivery) are delivered with fidelity (RQ1)	The in-person training was delivered as planned. While cascading took place in most schools, it was not consistently delivered as intended. Attendance at the two regional online check-ins was mixed, although nearly all schools had at least one representative at each session. Delivery of the regional online check-in meetings deviated slightly from the delivery team's original plan. While implementation support visits were largely delivered with fidelity, this was not the case for one-to-one meetings. Classroom sessions were delivered regularly, but implementation was uneven across schools and teachers. Participants did not consistently adhere to the prescribed classroom session length or all core components.
Intervention components (training/ongoing support/training materials/delivery resources) are of appropriate quality (RQ2)	Lead teachers reported high levels in both understanding and confidence following initial training from BRS. The training was praised for the accessibility of the delivery and the opportunity to practice key strategies. Some lead teachers explained that they enjoyed the opportunity to learn from other schools during the regional check-in meetings, while others would have preferred a more interactive format. Year 7 teachers' views of available resources for teaching problem-solving improved substantially after they had accessed the programme materials for TM, which suggests that the materials were useful. Participants' views about the problems varied, with some praising the selected problems for encouraging pupils to think deeply, while others believed the problems could be too easy for high attainers or too difficult for pupils less secure in maths. While there was no indication that the programme was too long, some participants wanted the programme to run for longer to sustain practice and to continue a routine. Some Year 7 teachers also wanted more flexibility regarding the timings because they believed the duration of sessions was dependent on pupil ability.
Resources/training/ongoing support are appropriate for schools with above-average proportion of FSM pupils (RQ3)	Survey and qualitative findings provide no indication that training, resources, and materials were experienced differently by schools with above-average numbers of FSM eligible pupils.

Barriers to delivery do not impede the programme being delivered with fidelity (RQ4)

Lead teachers praised the clarity of delivery and the trainers' expertise and ability to answer questions. Lead teachers explained that the cascading resources (the cascading video and PowerPoint) were comprehensive, making preparation straightforward. Time constraints impacted lead teachers' ability to cascade the training in their school, with some schools unable to dedicate an hour to cascade the training. Time could also be a barrier to classroom delivery, with participants noting that the problems had varying levels of difficulty which impacted the length of time spent on each session.

Evidence of promise

The following section explores questions around evidence of promise, covering research questions 6 to 11. The section focuses on the pathways to improvements laid out in the ToC (Figure 3) with a focus on proximal (short-term) outcomes. As this was a pilot evaluation involving 28 schools taking part in a 20-week programme, quantifying impact at the teacher or pupil level was beyond the scope of the study.

RQ6: Is there evidence that teachers are motivated to engage with the programme and change their classroom practice?

Key findings

Year 7 teachers and lead teachers were motivated to engage with the programme. Their engagement was helped by pupils' engagement, demonstrating how pupil motivation can shape teachers' experiences of a programme. Year 7 teachers reported being less motivated to engage with the programme where pupils struggled with problem-solving, or when they covered problems with their class which they did not deem suitable for problem-solving.

The pre-delivery survey indicated a need for additional support with problem-solving. While 85 out of 129 (66%) lead teachers and Year 7 teachers reported having some resources available for teaching problem-solving (Appendix 11 Table 29), only 5 out of 85 (6%) respondents found them very useful (Appendix 11 Table 30). Similarly, lead teachers and Year 7 teachers reported that department-wide guidelines for teaching problem-solving were limited, with just 39 out of 129 (30%) reporting their existence (Appendix 11 Table 31). Moreover, only 26 out of 129 (20%) lead teachers and Year 7 teachers reported that approaches to problem-solving were 'consistent' or 'very consistent' across teachers at their school, underlining the need for a more consistent approach to problem-solving across pilot schools (Appendix 11 Table 32).

Year 7 teachers and lead teachers were motivated to engage with the programme and saw value in improving their classroom practice, although enthusiasm declined over time. In the post-delivery survey, 68 out of 85 (80%) of all lead teachers and Year 7 teachers reported that they had felt 'enthusiastic' or 'very enthusiastic' about delivering the programme at the start of the programme. This figure dropped slightly to 58 out of 84 (69%) of all lead teachers and Year 7 teachers who said they felt that way by the end. Lead teachers were more likely than Year 7 teachers to begin the programme feeling very enthusiastic ($p = 0.048$, Appendix 11 Table 33), most likely driven by their more detailed knowledge of the programme and appreciation of why it was needed at their school, particularly if they signed the school up to participate in the pilot. By contrast, teachers of classes where many pupils were less secure with maths were more likely to begin with only slight enthusiasm ($p = 0.014$), potentially reflecting concerns about how well the materials would fit their pupils' needs (Appendix 11 Table 34). However, given the small sample size, this finding must be interpreted cautiously.

Interviews with lead teachers and focus groups with Year 7 teachers suggest a key driver for their motivation was the perception that the programme enabled pupils to think about maths in practical and applied ways. This was seen as different to 'business as usual' maths teaching. Moreover, lead teachers and Year 7 teachers saw the programme as enabling teachers and pupils to build strong relationships (see RQ5), facilitated by teachers engaging with pupils in a sustained dialogue about process rather than correcting pupils and focusing on the right answer.

Lead teachers and Year 7 teachers' engagement was further helped by pupils' engagement, demonstrating how pupil motivation can shape teachers' experiences of a programme. There was a view that pupils were motivated because they felt able to take risks and were less concerned about the consequences of making mistakes (see RQ9). Lead teachers and Year 7 teachers believed that pupils were more open to participating because they understood that the sessions were about articulating the thinking process rather than having the right answer. This resulted in sustained interactions with pupils that Year 7 teachers enjoyed.

The teachers quite enjoyed going and having...conversations with the students and actually having a bit of back and forth and playing devil's advocate with the students. In a normal lesson...it's nowhere near this level...It was really good for rapport building. (Lead teacher—interview)

Year 7 teachers reported being less motivated to engage with the programme where pupils struggled with problem-solving. For instance, a teacher explained during a focus group that TM sessions were not their 'favourite' because those pupils that were already less secure in maths struggled with the question and became disengaged as a result.

It's not my favourite thing to do with them because they're going to switch off very quickly. (Year 7 teacher—focus group)

Year 7 teacher motivation was also affected by their views on the particular problems. For instance, a lead teacher reflected that the Year 7 teachers had enjoyed the problems more at the beginning, as they regarded them to be more suitable to problem-solving; in contrast, the Year 7 teachers regarded the questions towards the latter stages of the programme as being technical and requiring certain mathematical knowledge which some pupils did not have. It is therefore possible that the slight decline in lead teachers' and Year 7 teachers' enthusiasm reported in comparisons between the pre- and post-delivery survey is partly driven by their views about the suitability of the weekly problem for their pupils and/or problem-solving.

RQ7: How effective is the train-the-trainer²⁷ model in supporting programme objectives?

Key findings

The initial training was effective in building preparedness among lead teachers, although Year 7 teachers seemed less prepared about delivery following cascading than lead teachers. A key reason impacting Year 7 teachers' confidence related to cascading sessions was from not providing Year 7 teachers with the opportunity to rehearse key session elements.

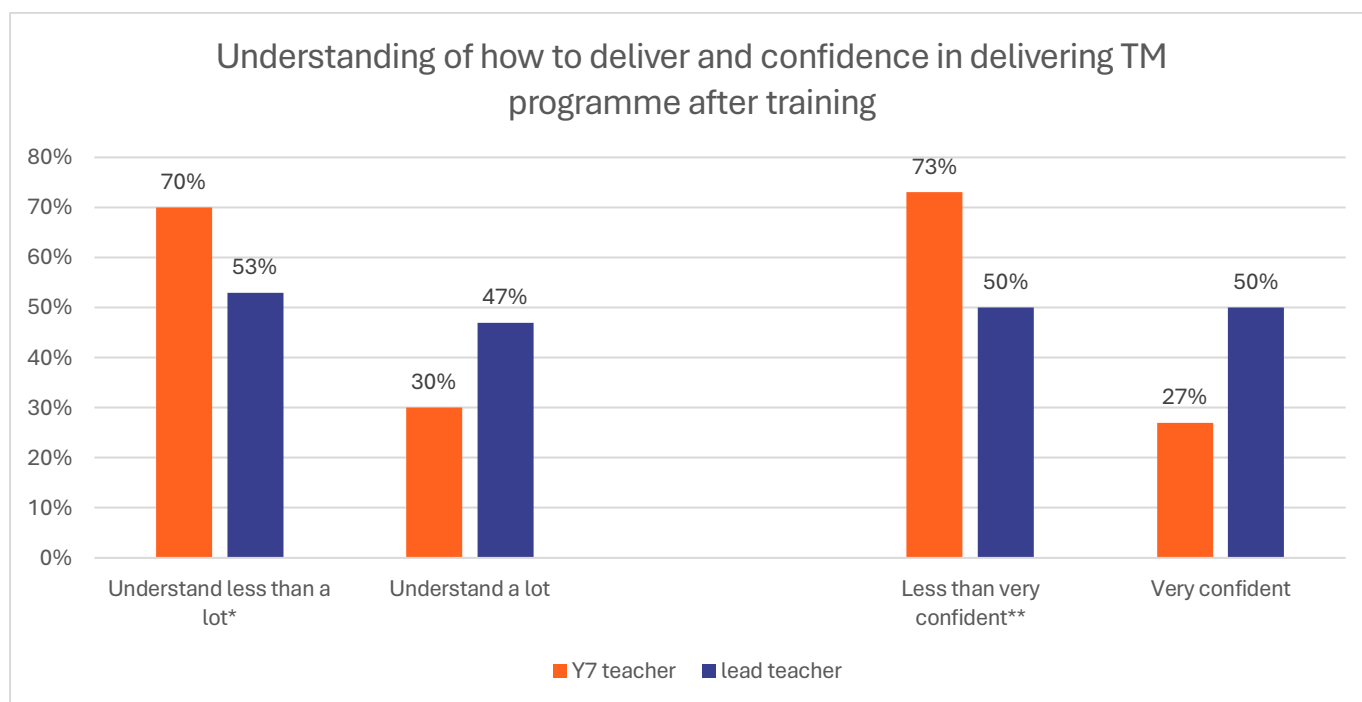
Evidence from the post-delivery survey suggests that the initial training was effective in building preparedness among lead teachers. Nearly all lead teachers reported feeling 'confident' or 'very confident' about cascading the training to their colleagues after receiving training from BRS, with no significant differences across teacher characteristics.²⁸

On the other hand, **Year 7 teachers seemed less prepared about delivery following cascading than lead teachers, suggesting that the cascading model had a limited effect for teachers.** Only 13 out of 44 (30%) reported understanding 'a lot' about how to deliver the TM programme after cascading, compared to 14 out of 30 (47%) of the lead teachers after their training with BRS. Similarly, 15 out of 30 (50%) lead teachers reported feeling 'very confident' about delivering the TM programme in their classrooms after their training with BRS, whereas only 12 out of 44 (27%) Year 7 teachers reported feeling the same after cascading (Figure 9, Appendix 11 Table 35).

²⁷ The train-the-trainer model refers to the model whereby *only* lead teachers attend the training and then share or 'cascade' the learning with the teaching staff participating in the pilot.

²⁸ Exact figures have been suppressed to reduce the risk of disclosure.

Figure 9: Understanding of and confidence in delivering TM programme after training by role



Base: All respondents of the post-delivery survey who reported attending a training session (Total, $N = 74$, Y7 teachers, $n = 44$, lead teachers, $n = 30$)

Source: Post-delivery survey

*This includes the options 'Not at all', 'Somewhat', or 'Quite a bit'. Categories have been combined to avoid the risk of disclosure.

**This includes the options 'Not confident at all', 'Only slightly confident', or 'Confident'. Categories have been combined to avoid the risk of disclosure.

One view from Year 7 teachers was that being 'daunted' was a common feeling when adopting a new classroom practice, although the internal training somewhat helped in relieving this concern.

You felt a bit daunted by it, but actually once you got in there and you got teaching with it, it was easy enough...It would have been more daunting without the session. (Year 7 teacher—focus group)

Year 7 teachers who felt confident following delivery attributed feeling confident and prepared to the clear and self-explanatory resources used during cascading (for example, the video) as well as the clarity around the messaging from their lead teacher, which they explained was to focus on the *process* of arriving at the answer rather than whether the answer is right.

A key facilitator to teachers feeling confident about delivery was previous experience with teaching problem-solving.

For instance, a Year 7 teacher explained that they had already been used to asking questions to encourage pupils to plan, monitor, and evaluate their proposed approach for answering a mathematical problem, so were less 'fazed'. For them, the cascading session simply helped to formalise and provide structure to what they had already been doing in their classroom practice.

Year 7 teachers who noted during focus groups that they lacked confidence and knowledge about the programme cited two reasons for this. First, there was a view that the time gap between the cascading session and the first classroom session had been too long, which resulted in Year 7 teachers forgetting key elements by the time they started delivering the session. However, it was unclear how long the reported time gap was where this was flagged as a challenge. A second reason related to cascading sessions not providing Year 7 teachers with the opportunity to apply some of the learning via a practical; this resulted in Year 7 teachers struggling to imagine what a session would look like in practice.

For us as a team, it would have been nice to have a bit of role-play beforehand, but then after the first session, it was fine. (Year 7 teacher—focus group)

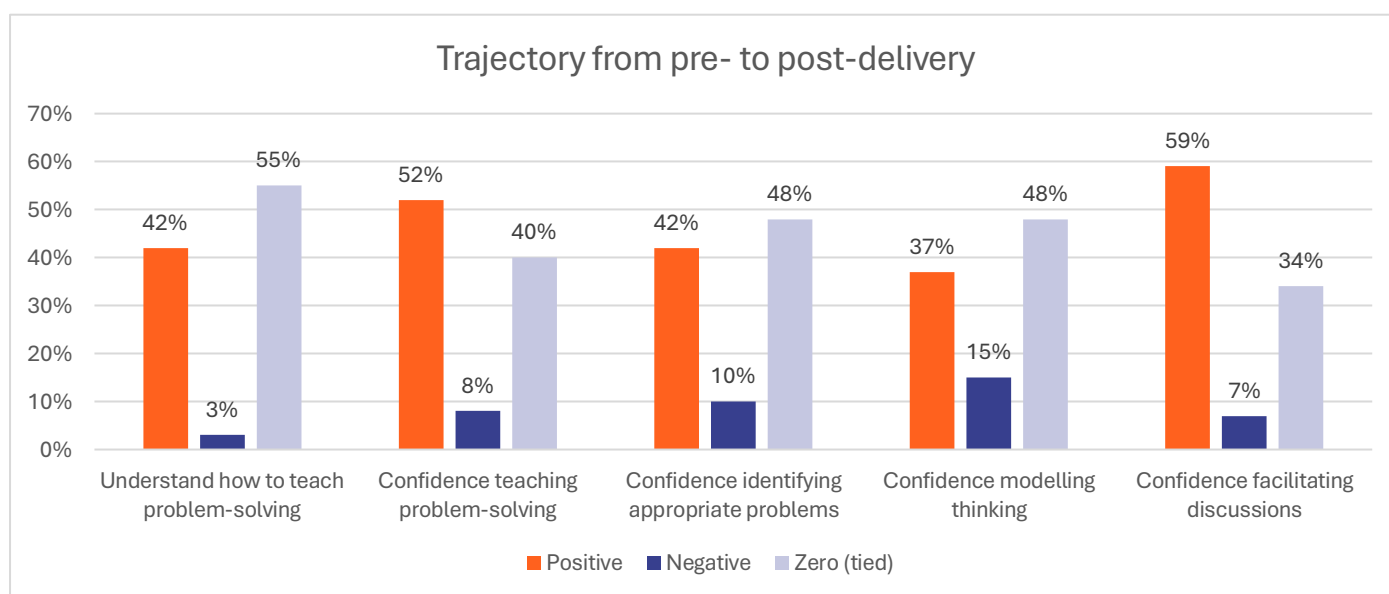
RQ8: Is there a change in teachers' confidence and knowledge in teaching the tackling of unfamiliar mathematical problems?**Key findings**

Pre-post analysis shows positive shifts in teachers' understanding of and confidence in teaching problem-solving. However, when asked specifically about benefits resulting from the programme, few lead teachers and Year 7 teachers reported clear personal gains. Those lead teachers who identified improved confidence in interviews noted that the routine of teaching problem-solving following the same structure helped develop confidence.

Survey evidence indicates overall positive shifts in understanding of and confidence in teaching problem-solving, though mixed patterns emerged in confidence around other aspects of problem-solving. Pre-post analysis of questions in the pre-delivery and post-delivery surveys confirmed improvements in lead teachers' and Year 7 teachers' self-reported understanding, confidence in teaching problem-solving, and confidence in facilitating discussions. The trajectory data in Figure 10 shows that the largest positive shift was in teachers' confidence facilitating pupil discussions (36 out of 61, 59% reporting improvement), followed by confidence teaching problem-solving (32 out of 62, 52%). The highest negative shift was in confidence modelling their own thinking, where 9 out of 62 (15%) reported a decline. Similarly, 6 out of 62 (10%) reported a decline in confidence in identifying appropriate problems. As previously mentioned (see Data analysis), the limited response rate means that the findings need to be treated with caution.

The negative shifts could have been an indication of reduced confidence, but when teachers were explicitly asked if reduced confidence was a negative outcome of the programme in the post-delivery survey, none of the teachers reported this. This pattern suggests that the negative patterns may instead be explained by teachers realising previously unrecognised gaps in their practice, correcting an initial overestimation once they better understood 'what they didn't know' through participation in the programme. At the same time, a substantial proportion of teachers' answers indicated no change (tied responses), particularly for understanding how to teach problem-solving (34 out of 62, 55%) and confidence modelling thinking (30 out of 62, 48%) (Appendix 11 Table 36).

Figure 10: Teacher understanding and confidence trajectory pre- to post-delivery



Base: Respondents who completed both pre- and post-delivery survey ($N = 62$)

Source: Pre- and post-delivery survey

Relatively few teachers reported clear personal gains. The post-delivery survey indicates that a sizeable minority of teachers did not report improvements in their own understanding or confidence, with only 47 out of 85 (55%) survey respondents reporting improved understanding, while 55 out of 85 (65%) survey respondents reported increased

confidence as a positive consequence of participating in the programme (Appendix 11 Table 37). This may suggest that some teachers did not experience positive outcomes for themselves. In interviews with lead teachers and focus groups with Year 7 teachers, participants tended to be reluctant to discuss outcomes *for themselves*, focusing primarily on benefits *for pupils*. This reluctance to discuss their own outcomes may help explain the less positive survey findings. During the learning workshop, a delivery team member explained that some participants' reluctance to note positive outcomes from programme participation could have been connected to the delivery team not placing emphasis on the benefits teachers could experience from the programme.

In the training sessions about the impact it's going to have on the young people, I'm not sure that we are as explicit about the impact that it's going to have on the teachers. (Delivery team member—learning workshop)

It is possible that a stronger emphasis on teacher impact could have resulted in Year 7 teachers experiencing the cascading sessions as something directly relevant to their classroom practice and potentially increasing their motivation to engage with the programme.

Where lead teachers and Year 7 teachers explained that their confidence improved during interviews and focus groups, they attributed this to the resources, notably the script and prompts. These reminded them to consistently 'think aloud' and focus on the importance of the process of arriving at an answer rather than the answer itself.

It reminds us to use that metacognitive talking out loud...it's that thinking out loud that I think then really benefits the students. (Year 7 teacher—focus group)

Those lead teachers who identified improved confidence in interviews noted that the routine of teaching problem-solving following the same structure helped develop confidence. This was echoed by the delivery team, who explained in an interview that the consistent delivery structure enabled teachers to apply strategies across year groups even without using the same problems.

Indeed, lead teachers and Year 7 teachers reported in interviews and focus groups that they started using TM strategies with other year groups, such as Year 8, Year 9, and even Year 11 pupils, which is a positive unintended consequence outlined in the ToC (see RQ5). The decision to embed problem-solving strategies in other year groups was driven by lead teachers and Year 7 teachers feeling more confident and knowledgeable about teaching problem-solving and seeing the benefits of the strategies in their own practice.

It's not just transformed my Year 7 lessons, but it's transformed...my teaching as well. (Lead teacher—interview)

Lead teachers and Year 7 teachers provided some concrete examples during interviews and focus groups what their changed practice looked like as a result of the benefits they experienced through participating in the programme. These included:

- **Becoming more reflective practitioners**—for instance, a Year 7 teacher explained that as a result of encouraging pupils to use different approaches to solving problems, they had become more open themselves to using and trying out different methods to teach mathematical concepts.
- **Being more pupil-led**—there was a view among Year 7 teachers that the emphasis on the debrief led them to do more pupil-led learning, such as getting pupils to discuss among themselves and self-correct.
- **Asking more meaningful questions**—for instance, a lead teacher explained that they started asking pupils to reflect on whether their method was working rather than telling them in their classroom practice; they identified exposure and usage of different mathematical problems as beneficial in helping them formulate different types of questions for pupils.

RQ9: Is there evidence that pupils are motivated to engage in the programme?

Key findings

Pupils were generally motivated to engage with the programme. Lead teachers, Year 7 teachers, and pupils identified several drivers for pupil engagement, including problem-solving being different to business as usual. Pupil ability shaped pupil engagement, with examples of pupils less secure in maths or high-ability pupils demonstrating lower levels of engagement.

Findings from the post-delivery survey indicate that pupils were generally motivated to engage with the programme, with moderately high levels of reported engagement sustained throughout delivery. In the post-delivery survey, 64 out of 85 (76%) Year 7 teachers and lead teachers said their pupils were ‘engaged’ or ‘very engaged’ at the beginning of the programme, and this proportion remained stable at 67 out of 85 (79%) by the end of the programme (Appendix 11 Table 38).

Interviews with lead teachers and Year 7 teachers provided several examples to illustrate pupil engagement, including pupils regularly reminding teachers about problem-solving lessons or asking whether they ‘are doing problem-solving today’.

There's normally quite a buzz... they'll be like, are we doing problem-solving today? Are we going to do the problem? Oh, no, we're doing it tomorrow. And there's disappointment if I say [we're] not doing it today. (Lead teacher—interview)

Lead teachers, Year 7 teachers, and pupils identified several drivers for pupil engagement. There was a view that the problems were practical and ‘real’ rather than abstract. Relatedly, lead teachers, Year 7 teachers, and pupils collectively identified problem-solving as different to business as usual, for instance, through teachers emphasising the value of different methods and a focus on process rather than getting the right answer. Pupils also enjoyed being exposed to their peers’ thought processes, as this was not something they had been accustomed to in their previous maths lessons.

Different people have different methods of working it out, and it's interesting to find out how other people's brains work and how it like, how they see the question when they first hear it. (Pupil—interview)

Interviews and focus groups with lead teachers and Year 7 teachers highlighted that pupil ability could affect pupil engagement. There was a view that pupils who were less secure in maths could become disengaged due to the literacy demands—such as long questions for some problems—and unfamiliarity with mathematical terms; as a result, they were more likely to become overwhelmed or disrupt the session.

Lead teachers and Year 7 teachers pointed out that high-ability pupils could also demonstrate lower levels of engagement, particularly where they found the questions too easy. This resulted in teachers introducing extension tasks to keep them engaged, such as asking pupils to create their own mathematical problem-solving questions.

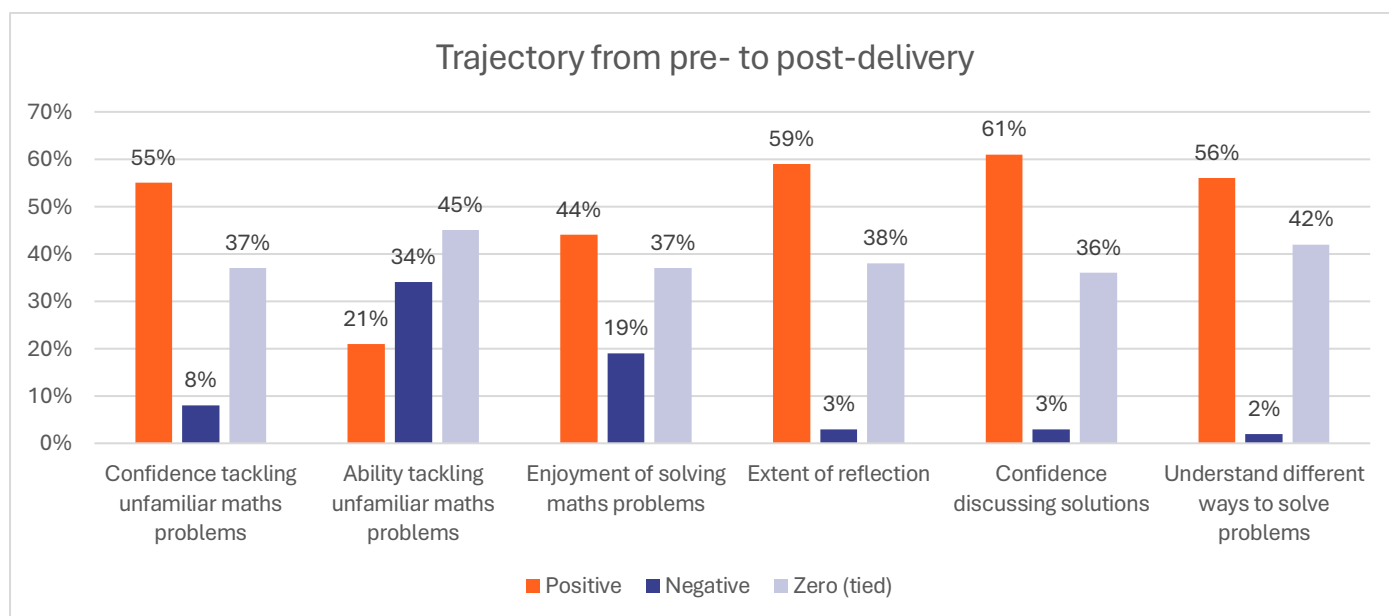
Observations of classroom delivery indicated that—as with most lessons—pupil motivation was not consistent across all sessions and/or pupils. For instance, during one observation, several pupils finished the independent problem-solving well before time and became distracted. Similarly, when asking pupils to complete a mathematical problem and talk us through their approach, some pupils appeared more engaged in the activity than their partner. While there were no instances of pupils not sharing their responses and process of arriving at an answer, it is important to recognise that the children who participated in the paired interview had been selected by their teacher and were possibly more engaged in the programme than some of their peers (see Limitations).

RQ10: Is there a (perceived) change in pupils' approaches to unfamiliar mathematical problems?**Key findings**

There were perceived improvements in pupil confidence to solve mathematical problems or confidence in discussing problem-solving. The majority of post-delivery survey respondents explicitly identified 'improved problem-solving' as a positive outcome of the programme. There was a view that the use of prompts contributed to pupils developing a problem-solving mindset.

Post-delivery survey²⁹ findings suggest that Year 7 teachers and lead teachers perceived some positive shifts in how pupils approached unfamiliar mathematical problems but also areas where change was less evident. There were perceived improvements from pre- to post-delivery in pupil confidence to solve mathematical problems ($p = 0.000$), the extent to which pupils reflect on how they approach a problem ($p = 0.000$), confidence in discussing problem-solving ($p = 0.000$), and understanding that there can be multiple solutions to a problem ($p = 0.000$). However, these gains did not extend to pupils' reported improvement in ability ($p = 0.2295$) or increase in enjoyment of problem-solving ($p = 0.0237$) (Appendix 11 Table 39). Figure 11 shows the trajectory from pre- to post-delivery for pupil ability, confidence, and enjoyment as reported by lead teachers and Year 7 teachers. However, given the small sample size, this finding must be interpreted cautiously.

Figure 11: Pupil confidence, ability, and enjoyment trajectory pre- to post-delivery



Base: Respondents who completed both pre- and post-delivery survey ($N = 62$)

Source: Pre- and post-delivery survey

²⁹ To assess changes in pupils' approaches to unfamiliar mathematical problems, teachers were asked to rate the average Year 7 pupil they taught across six dimensions; confidence to tackle unfamiliar mathematical problems, ability to tackle unfamiliar mathematical problems, enjoyment of solving mathematical problems, the extent to which pupils reflect on how they approached a problem once they had come up with a solution, confidence to discuss mathematical problem-solving with their peers, and the extent to which pupils understand that there can be more than one way to solve a mathematical problem. For confidence, ability, and discussion, the response options were 'not at all confident', 'only slightly confident', 'confident', and 'very confident'. For enjoyment and reflection, the options were 'not at all', 'somewhat', 'quite a bit', and 'a lot'. For whether pupils understood that problems could be solved in different ways, the options were 'none of my students know this', 'only a few of my students know this', 'most of my students know this', and 'all of my students know this'.

Interviews with lead teachers and focus groups with Year 7 teachers suggested that pupil confidence manifested itself in different ways. This included pupils being more confident in ‘having a go’, in making mistakes, and in speaking out loud in class and sharing their approach with peers.

A key perceived driver of improvements in pupils’ confidence was the emphasis on process, which lead teachers and Year 7 teachers believed was encapsulated by the debrief, where pupils shared their approach with the rest of the class. A lead teacher explained that their pupils had initially been reluctant but had grown in confidence over time.

When it started, they were very hesitant to speak out loud to the whole class in case they were wrong...whereas now they're much more confident and happy. (Lead teacher—interview)

Lead teachers and Year 7 teachers who observed improvements in pupils’ confidence sometimes used confidence and resilience—characterised by a perseverance and unwillingness to give up—interchangeably. There was a view that resilience was an essential attribute to build and develop so that pupils were in a position to successfully tackle exams (see EEF, 2018; Crenna-Jennings, 2018).

My thoughts are that when they get to a GCSE question [...] if they've got no resilience, they'll just, won't even try they won't even do anything, so they definitely need to build that up so that they are just prepared to have a go and it definitely is helping the Year 7s with that. (Lead teacher—interview)

Around 59 out of 85 (69%) post-delivery survey respondents explicitly identified ‘improved problem-solving’ as a positive outcome of the programme (Appendix 11 Table 40). It is possible that post-survey delivery respondents who did not identify ‘improved problem-solving’ as an improved outcome did so because of challenges they identified with the accessibility of problems for pupils less secure in maths and/or the lack of challenge some problems presented to high-ability pupils (see RQ2, RQ4, RQ9). An alternative explanation is that they simply did not believe it led to pupils’ improved problem-solving.

In focus groups and qualitative interviews with lead teachers and Year 7 teachers, pupils’ improved problem-solving was reflected in lead teachers and Year 7 teachers explaining how **pupils were more systematic in approaching mathematical problems**, characterised by carefully and critically engaging and dissecting the question first before attempting an answer. There was a view that the use of prompts contributed to pupils developing a problem-solving mindset, where pupils moved from wanting to be the first to answer a question to learning how to plan, monitor, and evaluate their work.

They're just trying to think more critically and work in a more systematic way to get to any solutions. (Year 7 teacher—focus group)

Although interviews with lead teachers and Year 7 teachers highlighted challenges around the accessibility of problems for pupils who were less secure in maths or EAL pupils (see RQ2), there was no indication in the qualitative data that certain groups of pupils benefitted more or less than others.

RQ11: Does the programme appear to be sustainable?

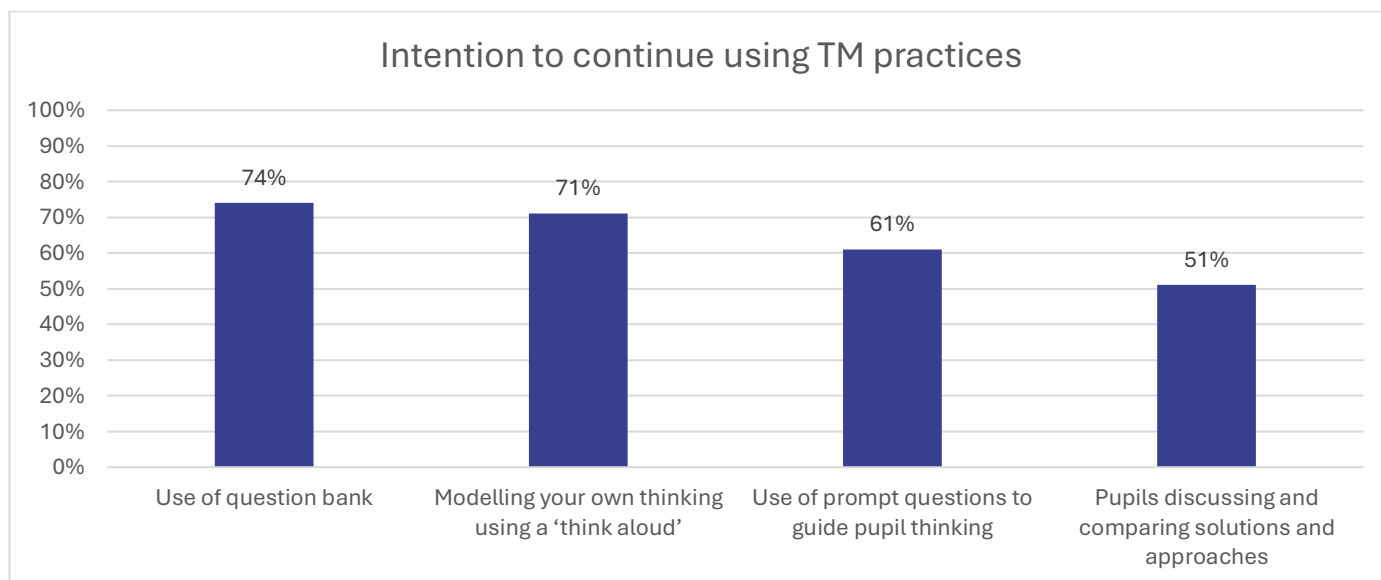
Key findings

Lead teachers and Year 7 teachers were motivated to continue many of the practices introduced through the programme, notably the use of the question bank and modelling their own thinking via a ‘think aloud’. One reason for continuing TM practices was to ensure pupils were suitably prepared for their GCSEs, which assess mathematical problem-solving. Pre-and-post-delivery survey comparisons also demonstrate a more consistent departmental approach to teaching problem-solving, pointing to TM practices becoming more embedded within schools’ maths departments.

Post-delivery survey responses indicate that teachers were motivated to continue many of the practices introduced through the programme. 63 out of 85 (74%) Year 7 and lead teachers reported that they wanted to keep using the TM question bank. Similarly, 60 out of 85 (71%) planned on continuing to model their own thinking through a ‘think aloud’ (see Figure 12). Fewer respondents reported intending to continue using prompts to guide pupil thinking (52 out of 85, 61%) and

to sustain the practice of encouraging pupils to discuss and compare approaches (i.e., the debrief) (43 out of 85, 51%) (Appendix 11 Table 41). It is likely that this was due to some lead teachers and Year 7 teachers not seeing prompts as essential once pupils had become accustomed to the questions and a perception that some problems were unsuitable for some Year 7 pupils.

Figure 12: Intention to continue using TM practices



Base: All respondents of the post-delivery survey ($N = 85$)

Source: Pre- and post-delivery survey

In interviews and focus groups with **lead teachers and Year 7 teachers, they explained their willingness to continue with a range of practices, including using suitable problems, making use of prompt questions as well as carrying out the debrief.** Their responses centred mainly on continuing with the current cohort of Year 7 pupils. One perspective was that it was important for the cohort of Year 7 pupils to continue with key practices so that mathematical problem-solving would become 'ingrained'. This was regarded as important in the context of pupils' preparation for their GCSEs which assess problem-solving and require pupils to exhibit resilience. A different reason for continuing with TM practices was to maintain pupil engagement; for instance, a lead teacher explained that they had already started preparing a booklet for the following academic year, as pupils had become accustomed to problem-solving.

We've got a problem-solving starter book in Year 8 in preparation for the fact that, okay, if Year 7s are used to doing this, we're going to need something in Year 8. (Lead teacher—interview)

Post-survey delivery indicates that teacher confidence plays a role in sustaining these practices. Lead teachers and Year 7 teachers who reported feeling 'confident' or 'very confident' teaching problem-solving were significantly more likely to say they would continue using prompts ($p = 0.003$, Appendix 11 Table 42). Among those who felt 'confident' or 'very confident', 24 out of 29 teachers (83%) reported intending to continue using prompts, compared with 28 out of 56 teachers (50%) among those who reported feeling 'not at all confident' or 'only slightly confident'. This indicates that embedding practices may be partly contingent on teachers' confidence with teaching problem-solving. However, given the small sample size, this finding must be interpreted cautiously.

Alongside lead teachers' and Year 7 teachers' commitment to continue with TM practices, our findings suggest that there is a more consistent departmental approach to teaching problem-solving. Among Year 7 teachers and lead teachers who responded to both the pre-delivery and post-delivery survey, 24 of the 39 (62%) who initially reported having 'not at all' or 'only slightly' consistent approaches to teaching problem-solving at their school moved to reporting 'consistent' or 'very consistent' approaches ($p = 0.008$, Appendix 11 Table 43). These improvements suggest a strengthening of shared practice across teachers and a coherent, embedded approach that may be an indication of sustainability. However, given the small sample size, this finding must be interpreted cautiously.

Evidence of success

Table 12 includes a summary of the evidence of success presented for the success indicators related to evidence of promise, covering RQs 6–11.

Table 12: Success indicators related to evidence of promise

Success indicators	Evidence of success
Teachers are motivated to engage in the programme and change classroom practice (RQ6)	Year 7 teachers and lead teachers were motivated to engage with the programme and saw value in improving their classroom practice, although enthusiasm declined over time. A key driver for their motivation was the perception that the programme enabled pupils to think about maths in a practical and applied way. Participants' engagement was helped by pupils' engagement, demonstrating how pupil motivation can shape teachers' experiences of a programme. Year 7 teacher motivation, however, was affected by some problems being less suitable to problem-solving than others; for instance, certain problems did not lend themselves to a debrief.
The train-the-trainer (TTT) approach results in lead trainers feeling confident in supporting the Year 7 teaching team (RQ7) The TTT approach results in practitioners feeling knowledgeable about key elements of the programme (RQ7)	While the initial training was effective in building preparedness among lead teachers, Year 7 teachers seemed less prepared about delivery following cascading than lead teachers, suggesting that the cascading model had a limited effect for teachers. Factors impacting Year 7 teachers' confidence included what was perceived to be too long of a time gap between the cascading session and the first classroom session, resulting in Year 7 teachers forgetting key elements by the time they started delivering the session. A second reason related to cascading sessions not providing Year 7 teachers with the opportunity to rehearse key session elements.
Teachers are more knowledgeable and confident in teaching tackling of unfamiliar mathematical problems (RQ8)	Pre-post analysis shows positive shifts in teachers' understanding of and confidence in teaching problem-solving. However, when asked specifically about benefits resulting from the programme, few lead teachers and Year 7 teachers reported clear personal gains, potentially because they did not view the programme as an intervention aimed at changing teachers' practice. Those lead teachers who identified improved confidence in interviews noted that the routine of teaching problem-solving following the same structure helped develop confidence.
Pupils are motivated to engage in the programme (RQ9)	Pupils were generally motivated to engage with the programme. Lead teachers, Year 7 teachers, and pupils identified several drivers for pupil engagement, including problems being practical and problem-solving being different to business as usual. However, there was evidence that pupil ability could affect pupil engagement, with examples of pupils less secure in maths or high-ability pupils displaying lower levels of engagement in some sessions.
Positive change in pupils' approaches to tackling unfamiliar mathematical problems (RQ10) No evidence that disadvantaged pupils benefit less than their less disadvantaged peers (RQ10)	There were perceived improvements in pupil confidence to solve mathematical problems; the extent to which pupils reflect on how they approach a problem, confidence in discussing problem-solving, and understanding that there can be multiple solutions to a problem. Moreover, most post-delivery survey respondents explicitly identified 'improved problem-solving' as a positive outcome of the programme. However, it is worth highlighting that the evidence is indirect, with limited evidence from pupils about improvements directly related to mathematical problem-solving. There was no indication from the surveys or qualitative research that disadvantaged pupils benefitted less from the programme compared to their less disadvantaged peers.

Changed practice by school and teachers embedded throughout 20 weeks and intended to be sustained after 20 weeks (RQ11)

Participants were motivated to continue many of the practices introduced through the programme, notably the use of the question bank and modelling their own thinking via a ‘think aloud’. One reason for continuing TM practices was to ensure pupils were suitably prepared for their GCSEs, which assess mathematical problem-solving. Pre-and-post-delivery survey comparisons also demonstrate a more consistent departmental approach to teaching problem-solving, pointing to TM practices becoming more embedded within schools’ maths departments.

Readiness for trial

The following section explores questions around readiness for trial, covering RQs 12 to 15. The section draws on the interview with the delivery team, conversations with the delivery team during the learning workshop, as well as the evaluation team’s own reflections on the extent to which the programme is ready to be evaluated at a larger scale, based on findings from across all pilot research activities.

RQ12: Is the programme sufficiently manualised to allow delivery with fidelity to a larger number of schools?

Key findings

At present, the programme is partially manualised: comprehensive and high-quality materials exist but require additional guidance and clearer articulation of core components before delivering the programme to a larger number of schools. This includes, for instance, a more explicit distinction between non-negotiables and flexible elements.

A programme is sufficiently manualised if it enables a large number of schools to implement the programme as intended and without requiring additional support (see Hall *et al.*, 2025). At present, the programme is partially manualised: comprehensive and high-quality materials exist but require additional guidance and clearer articulation of core components before delivering the programme to a larger number of schools.

Documentation

The programme includes a comprehensive set of training, cascading, and classroom resources. A training handbook is provided to all training participants as well as all Year 7 teachers delivering the sessions and includes the programme timeline; a summary of resources; the purpose and format of the cascading session; the purpose and format of the delivery session, as well as a short description of the core components taking place during each intervention session, including the ‘think aloud’, ‘heuristics’ (prompts), and the ‘debrief’.

Cascading resources (cascading slides and video) mirror the content covered in the training and ensure—at least in theory—that those attending the training (lead teachers) receive the same key messages as those attending the cascading sessions (Year 7 teachers).

Classroom resources include a presentation and scripted teacher notes for each session. The notes explain to teachers how to approach the problem with their pupils as well as what kind of maths is required and what resources are needed; however, they do not include a breakdown of indicative timings; that is, how much time should be spent on each session activity. **It is possible that indicative timings could have helped more consistent delivery across schools.**

Core versus adaptable components

While the materials (training slides, training handbook, and cascading slides) clearly outline the core components of the programme, there is limited indication which session activities, if any, can be adapted to the local context. It is possible that a more explicit distinction between non-negotiables and flexible elements could have contributed to more schools delivering the programme as intended, for instance, adhering to the required time (see RQ1).

The delivery team distinguished in their materials between required and optional activities with cascading activities to ensure that all schools completed a minimum requirement of activities. However, they acknowledged in the learning

workshop that they could have been more explicit in the programme handbook about *what* they required schools to do in the session. It is possible that more prescriptive guidance beyond what was in the programme handbook would have led to less variation in the way cascading was delivered (see RQ1).

Monitoring and ongoing support

We cannot comment on how manualised the provision of ongoing support from the delivery team is at present, as we did not review any materials used by the delivery team during their visits. **It is possible, however, that some of the knowledge of how to conduct the visits (and other forms of ongoing support) is tacit**, as the delivery team currently consists of two members. If the programme is delivered at scale and involves more members of the delivery team providing ongoing support to schools, it will be essential for materials to exist to ensure there is minimal variation in the support provided to schools by the delivery team.

The delivery team highlighted during the learning workshop that they needed to make some changes to the way they captured data on in-school fidelity via lead teachers to ensure they had a better understanding of the extent to which schools adhered to core components (see RQ13).

Clarity of guidance

The materials are written in clear and non-technical language and are accessible. For instance, the handbook explains more technical terms (for example, heuristics) in simple language. The classroom slides are clear and concise, while the teacher notes adopt an informal, conversational tone, reflecting how a teacher would talk to their pupils. Lead teachers reported in interviews that the resources were accessible for non-subject specialists; this is particularly important in the context of delivering the programme to more schools, as it is possible that there would be a higher proportion of non-subject specialists delivering the sessions in a future trial.

RQ13: Are viable strategies in place to collect sufficient data to monitor compliance and fidelity?

Key findings

The delivery team collected data to monitor fidelity via delivery logs; however, they did not capture the extent to which all core components of the programme were delivered as intended, suggesting that the development of a more refined system for monitoring fidelity is an area of improvement for a future trial.

During implementation, monitoring of fidelity happened at two levels: first, the delivery team monitoring implementation in schools via the delivery log and implementation support visits; and second, lead teachers monitoring implementation in their schools via observations, discussions with staff, and internal forms.

The delivery team used the following principles in collecting data on fidelity:

- **Comprehensiveness**—sending the delivery log to both lead teachers to ensure a minimum of one log was sent back.
- **Brevity**—including few questions and limiting them largely to multiple choice.
- **Consistency**—being consistent with how they are distributed; for instance, the logs were sent out on Monday morning and followed by a reminder on Wednesday afternoon.

The delivery team reflected in a follow-up email in December 2025 that the switch from weekly to fortnightly delivery logs was beneficial, as it gave practitioners additional time to reflect on the changes to classroom practice being made.

Interviews with lead teachers confirmed that they appreciated the simple and familiar format.

Questions have been more or less the same, so we kind of know what to expect. (Lead teacher—interview)

Another view was that the repetitive format made lead teachers feel that submitting the logs was not the best use of their time, as they found themselves providing the same information every two weeks. This suggests that some lead teachers would be even less receptive to submitting weekly logs, as had been the case during the second roll-out.

Programmatic data from the delivery log reveals that submission of delivery logs was inconsistent, with delivery logs not submitted by some schools in several cycles. Submission rates fluctuated between 71% and 100% across the programme and were lowest in Week 8 and Week 10 (Appendix 11 Table 39) due to technical issues and school closures following heavy snowfall. Reflecting on the submission rates, the delivery team noted in a follow-up email correspondence in December 2025 that they generally considered submission rates high; however, they explained that comparisons with submissions rates during the second roll-out would make limited sense, as financial incentives offered to schools was linked to completion of monitoring and evaluation activities.

The delivery team members noted in their interviews that the data they collected to monitor fidelity was overly focused on gathering data on timings and whether all teachers delivered the session. This emphasis on timings and staff participation meant that the delivery team did not capture the extent to which other parts of the programme were being delivered, including the debrief. They acknowledged that this would have been an important component to collect data on during the pilot; the debrief tended to happen at the end of the session and was most at risk of not taking place due to teachers spending too long on other activities.

I think the other aspects kind of naturally happen. Yeah, the teacher introducing the problem, that's always going to happen. The children having a go at the problems... that's always going to happen. I think the debrief is the bit that can sometimes not happen, especially timing wise. (Delivery team member—interview)

Other than submitting the delivery logs, there was no prescribed or systematic way in which lead teachers monitored fidelity of delivery. For instance, the programme materials did not include fidelity checklists outlining where to be 'tight' and where to be 'loose' (EEF, 2024), which would have enabled lead teachers to systematically observe whether sessions were delivered as intended. As a result, their approaches to capturing fidelity of delivery in their school varied. Interviews with lead teachers highlighted two contrasting approaches to collecting data on delivery from teachers in order to determine whether the programme was implemented as intended:

- Formal approaches, such as discussing programme delivery as a standing agenda item, regularly observing teachers, and collating information about delivery via internal forms.
- Informal approaches characterised by informal lesson 'drop ins' and impromptu conversations with teachers.

RQ14: What outcome measures might be used in an efficacy trial?

Key findings

Progress Test in Maths (PTM) is an appropriate outcome measure for an efficacy trial. The problem-solving subscale raw score can be used as a secondary, exploratory outcome, as it aligns with the programme's focus on problem-solving. PTM is UK normed and offers practical whole-class administration with independent centralised scoring conducted by a test publisher. It has been extensively used in the EEF trials.

We carried out a structured review to identify an outcome measure suitable for capturing mathematical problem-solving for Year 7 pupils in a possible trial (see Appendix 12 for the detailed review methodology and findings).

We identified the Progress Test in Maths (PTM) as a suitable outcome measure. Based on the age group of pupils targeted by the programme, the PTM Level 11T would be appropriate as a baseline measure and the PTM Level 12 would be appropriate as an endline measure. The PTM is well-suited for an efficacy trial because it is UK normed on a large sample, providing reliable benchmarks for this population. It offers equivalent baseline and endline forms, comes in both digital and paper formats, and allows for whole-class administration, making it feasible to use at scale. However, there are some practical considerations to be mindful of in a trial: the assessment is 60 minutes in length, but would require additional time for setting up; this would need to be clearly communicated to schools in any recruitment materials, as this might exceed the time for typical lessons. Moreover, digital administration requires the use of earphones/headphones for parts of the assessment, which might have cost implications. A key strength of PTM is its strong validity and reliability, ensuring it provides robust and trustworthy evidence of pupil progress.

Although PTM is a suitable outcome measure for a trial, we proposed in our write-up of the desk-based review a small number of clarifications before a final decision is made about *how* PTM should be used (Appendix 12). This included a review of the scoring criteria for the more complex problem-solving items, and a check on the strength of evidence underpinning the ‘Understanding and applying mathematical processes through reasoning and problem-solving’ sub-score, including any factor analysis and indicative evidence on reliability and validity.

Independent centralised scoring conducted by the test publisher rather than class teachers and scoring of the process of problem-solving items strengthen PTM’s fit with the programme, even in the absence of a published detailed rubric. While a detailed scoring rubric for complex items is not available from GL Assessment, PTM is scored centrally by GL Assessment rather than by class teachers, providing a useful layer of blinding for trial outcomes. In addition, available materials indicate that several of the more demanding items award credit for the problem-solving process as well as final correctness, aligning closely with the programme’s emphasis on multiple approaches to problem-solving.

Evidence on subscales supports exploratory use alongside the use of the total score. We find that the PTM subscales have not undergone the same level of psychometric evaluation as the full test and are not designed to function as standalone measures. Nonetheless, PTM has been used extensively in the EEF evaluations (see Basharat *et al.*, 2023; Jay *et al.*, 2016), and raw subscale scores have been employed effectively as secondary outcomes alongside the full test score (see Demack *et al.*, 2022). PTM’s precursor, Progress in Maths (PiM), has also been used in many of the EEF trials. Correlations between the two versions are high ($r = .88$), suggesting strong continuity. This was one of the reasons that additional piloting of the measure was not deemed necessary.

Some evaluations have noted potential floor effects while using PTM, particularly with a higher level of the test (see Demack *et al.*, 2022) or when used with lower-attaining pupils (see Hodgen *et al.*, 2019). One trial has addressed this through the use of raw scores rather than age-standardised scores to enhance sensitivity at the lower end of the distribution (see Groom *et al.*, 2024).

Therefore, **we propose specifying the PTM total raw score as the primary outcome and the problem-solving subscale raw score as a secondary, exploratory outcome**, using raw-score analysis to maintain sensitivity. This aligns with the programme focus and provides a rigorous and practical basis for measuring pupil progress.

RQ15: Is the delivery approach appropriate to be delivered with fidelity at a larger scale?

Key findings

The intervention’s core components—the use of genuine problems; teachers modelling using a ‘think aloud’ approach; the use of prompts; and pupils discussing and comparing solutions and approaches can be delivered at a larger scale. However, the delivery team will need to consider more carefully what information to record from schools to learn more about the extent to which schools have adhered to the core components, thus enabling them to provide targeted support to those who do not deliver the programme with fidelity.

The intervention’s core components can be delivered at a larger scale. Key elements under consideration for the delivery team that relate to classroom delivery at scale include a) monitoring and b) dosage.

The delivery team noted that the delivery log used to monitor delivery required changes for a possible trial. While the logs were useful in helping the delivery team to identify schools that needed additional support, they did not capture all core components, including whether the debrief or the ‘think aloud’ were taking place across schools (see RQ13). Alongside capturing information on *all* core components to enable a full understanding of fidelity, the delivery team noted that it would be more efficient and consistent for the log to be completed by one lead teacher per school rather than two, as was intended during the pilot.

Regarding the session and programme duration, the delivery team emphasised that sessions should remain at 20 minutes to meaningfully cover the key components without affecting the curriculum. However, they were open to extending the duration of the programme; they noted that they had received feedback from pilot schools who decided to

continue the weekly sessions until the end of the academic year and had therefore decided to select their own weekly problems.

The delivery team's plans for what scaling to deliver in more schools would look like are clear and realistic, although there is variation in the level of detail and planning that has gone into each element. The delivery team's scaling plans centred on recruitment of schools, initial training, and ongoing support.

Recruitment of schools

The delivery team were confident they had enough capacity and the ability to recruit the required number of schools for an efficacy trial. They highlighted the need for more administrative support should the programme be delivered at scale but did not foresee this as a challenge. They formulated clear plans on how to market the programme and which schools they wanted to target. It might also be that the importance of problem-solving for GCSE maths, the limited offer of specific school-based interventions aimed at improving mathematical problem-solving skills, and the limited reach of the TM programme thus far will facilitate recruitment.

Marketing the programme

The delivery team identified the Research Schools Network³⁰ and the Maths Hubs³¹ as two recruitment pathways for marketing the programme and reaching the required number of schools. Given the national reach of the Research Schools Network and the Maths Hubs, they were confident that recruiting at scale was feasible with a) a longer timeline for recruitment compared to the pilot stage, and b) more internal capacity for administrative support. Using networks with a national reach will also be important to avoid a regional imbalance of participating schools, as this can have implications for the size of in-person training sessions (see RQ1 and further below).

Identifying target schools

The delivery team's emphasis on working with schools with higher-than-average proportions of FSM pupils, coupled with their mission of contributing to closing the attainment gap, meant that they were inclined to target schools in deprived areas for a possible trial. However, they noted that they did not want to make this a requirement, as there was a wider need for pupils to improve mathematical problem-solving and teacher practice.

The delivery team explained that they were actively considering recruiting multi-academy trusts (MATs) for a possible efficacy trial, as this had the advantage of being able to recruit several schools at once. They acknowledged though that recruiting MATs posed challenges, especially if schools within the trust had limited decision-making power and required sign-off from the MAT's central team to take part. From an evaluation standpoint, however, the inclusion of MATs presents challenges for randomisation, given the increased risk of contamination between schools within the same trust. This is something that evaluators will need to consider in the event of a future trial.

Initial training

The delivery team emphasised that the training model of delivering in-person training to two lead teachers was scalable without undermining implementation fidelity. They noted the importance of in-person training, as this was key for teacher engagement. They expressed the view that schools were more likely to commit to the programme if they allowed senior teachers to attend training for a day, highlighting the necessity of lead teachers being senior practitioners.

The delivery team understood that delivering the training to more than 50 schools required finding suitable training locations that were geographically accessible for schools to attend; to ensure a consistent experience of the initial training and make training viable, they were keen to ensure a minimum and maximum number of attendees per session.

³⁰ The Research Schools Network consists of a collaboration of 33 schools across England, supported by the Education Endowment Foundation.

³¹ The Maths Hubs programme is coordinated by the National Centre for Excellence in the Teaching of Mathematics (NCETM) and helps teachers, departments, and schools lead improvement in maths education, for instance by providing CPD opportunities. At present, there are 40 Maths Hubs in England.

They noted that a total of 40 attendees from 20 schools would be the upper limit. To avoid differences in group size, the delivery team will need to be mindful of recruiting ‘clusters’ of schools to minimise the risk of in-person training sessions being small (see RQ1 and further below) while ensuring that training locations are accessible for all participating schools.

Training more trainers

Training more schools before delivery in a possible trial would require the delivery team to expand and recruit additional staff. The delivery team explained that they would draw on a team of subject experts and evidence leads within BRS, who would have the required expertise and skills to deliver the training. To ensure they also had the required knowledge about the *programme*, including the evidence base, the materials, and the core components, the delivery team proposed trialling a dedicated three-day training programme for trainers. This is the typical training duration for trainers for school-based programmes at BRS.

Cascading

The delivery team wanted to retain a cascading model, as they deemed it ‘logistically impossible’ for all Year 7 teachers in addition to lead teachers to attend the initial training sessions. They acknowledged the importance of making changes so that cascading could be delivered consistently, thus increasing the likelihood of classroom sessions being delivered with fidelity. Suggested changes related to how best to communicate cascading to senior leaders and improving the resources available for the cascading session.

The delivery team suggested the MoU needed to be more explicit about cascading expectations, in particular, about the time required for cascading and what the time needed to be used for. They reflected that programmes work best in schools when senior leaders are clear on what is expected before signing up to a programme.

An idea under consideration at the time of writing was to provide Year 7 teachers (i.e., those teachers who had not attended the initial training sessions) with workbooks, which they could use during the cascading session. The delivery team explained that this would ensure that all teachers used the programme materials such as prompts more consistently in their sessions.

Producing teacher workbooks and then during that cascading session... the teachers have a kind of consistent workbook, and we find that when we have things like that, it's much more likely that those resources are going to be engaged with. (Delivery team member)

The delivery team also proposed including more video exemplifications as part of cascading sessions, alongside stipulating the need for teachers to rehearse key programme activities such as the ‘think aloud’ approach. They believed this had the benefit of bringing the programme ‘to life’ rather than keeping it ‘abstract’ for teachers.

A final adaptation concerned the removal of the optional elements, as they believed this undermined consistent cascading across schools. The delivery team were conscious that they needed to remove some of the optional elements if they ‘tightened’ expectations for cascading. They were conscious that they did not want to include too many activities to maintain teacher ‘buy-in’ and explained that schools tended to have existing mechanisms in place for ‘optional activities’ such as lesson drop ins.

[I]f we were to expect more in the kind of compulsory element, I think... we might have to eliminate the optional [elements]. [W]hat you will find is that lots of schools have mechanisms in place anyway for things like lesson drop ins...usually the head of department/second in department will be dropping into lessons and will get a good sense of what's going on. (Delivery team member)

Regional online check-ins

The delivery team thought that regional online check-in meetings were scalable; the only plan under consideration was to enhance the interactivity of the session by allowing more collaboration between schools, for instance, by assigning schools to breakout rooms. A question for BRS to consider is whether to have an online meeting for all clusters or whether to deliver several check-in meetings to different regional clusters, which mirrors the approach to in-person training.

Online meetings

It was evident to the delivery team that individual online meetings with each school were not scalable. They suggested online meetings could be a mechanism to monitor in-school fidelity, for instance, by selecting a random sample of schools to reach out to. However, at the time of the learning workshop in July 2025, the delivery team’s plan for this had not yet been refined.

Implementation support visit

The delivery team believed that the implementation support visit would continue to serve as an important function to support schools struggling with the programme and to maximise fidelity in a possible trial. While they highlighted the logistical challenge of visiting large numbers of schools, they noted that trainers could support with implementation support visits so that a proportion of schools (around 20%) could be visited. They suggested that visiting the same proportion of schools in a trial would be feasible with a larger pool of trainers. However, the delivery team also suggested that implementation support visits could be virtual, although this would make it more difficult for sessions to be observed. A suggestion to resolve this challenge was that lead teachers could be asked to observe several sessions and then share their reflections with the delivery team in writing before an online meeting.

Table 13 includes a summary of the evidence of success presented for the success indicators related to readiness for trial, covering RQs 12–15.

Table 13: Success indicators related to readiness for trial

Success indicators	Evidence of success
Core components identified for key programme activities (RQ12)	The programme is partially manualised: comprehensive and high-quality materials exist but require additional guidance and clearer articulation of core components before delivering the programme to a larger number of schools. This includes, for instance, a more explicit distinction between non-negotiables and flexible elements.
Sufficient data collected to monitor compliance and fidelity (RQ13)	The delivery team collected data to monitor fidelity via the delivery logs; however, they did not capture the extent to which all core components of the programme were delivered as intended, suggesting that the development of a more refined system for monitoring fidelity is an area of improvement for a future trial.
Time spent on data entry (for example, completing delivery logs) is proportionate to its purpose (RQ13)	There is no indication from the qualitative interviews that time spent on the delivery logs was burdensome.
Outcome measure(s) identified that is/are <ul style="list-style-type: none"> a) aligned to intended outcomes; b) able to measure changes that we expect to see; and c) convenient to administer in secondary schools (RQ14) 	The Progress Test in Maths (PTM) has been identified as a suitable outcome measure that meets the criteria. It aligns with the programme’s intended outcomes through broad mathematical coverage and a specific problem-solving subscale that reflects the programme’s focus of tackling non-routine mathematical problems. It can detect change over time via parallel baseline/endline forms and the use of raw scores to maximise sensitivity (particularly at the lower end) as evidenced through its wide usage in other educational trials, including for the EEF. It can therefore be used comparatively. It is convenient to administer in secondary schools as it is well known and widely used by schools and teachers. It also offers practical whole-class delivery and independent centralised scoring conducted by the test publisher rather than class teachers which reduces the burden of administration on schools.
Clear and realistic plan for what scaling to deliver in more schools would look like and require (RQ15)	The delivery team’s plans about what scaling to deliver in more schools would look like and require are well explicated, although not refined at the time of writing. There is recognition that administrative support would be required to recruit at scale and that the delivery team would need to leverage their networks to reach more schools. The team is also clear that they would need to

Plan for scaling would entail little or no modification to content of in-school delivery (RQ15)

train more delivery staff to deliver at scale, especially given that in-person implementation support visits are a key part of the programme.

Plans for scaling would require changes to the cascading model, which is delivered in schools by lead teachers.

Assessment of success

Table 14 includes an assessment of the success indicators co-developed with the delivery team at the beginning of the pilot (see Marshall *et al.*, 2024). The assessment is based on the evidence presented against each success indicator (see Table 11, Table 12, Table 13) where the success indicator is not met, a short explanation is provided to explain how the challenge might be addressed.

Table 14: Assessment of success indicators

Pilot criteria	Success indicators	Assessment of success
Feasibility of implementation	Programme activities (training/ongoing support/in-class delivery) are delivered with fidelity (RQ1)	Somewhat—challenge can be addressed but will require careful and detailed planning. A main challenge will be to make changes to cascading so that cascading sessions can be delivered more consistently across schools. Changes under consideration include communicating the requirements of cascading more clearly to senior leaders and improving the resources available for the cascading session. Another challenge for the delivery team will be to refine their systems for monitoring delivery so that settings who do not adhere to the core components receive targeted and timely support.
	Intervention components (training/ongoing support/training materials/delivery resources) are of appropriate quality (RQ2)	Somewhat—challenge can be addressed but will require careful and detailed planning. A key challenge will be to make problems more accessible by including different difficulty levels for the same questions. The programme would also benefit from a review of the sequence of questions so that easier questions are covered at the beginning of the programme, as well as ensuring that selected questions lend themselves to a debrief.
	Resources/training/ongoing support are appropriate for schools with above-average proportion of FSM pupils (RQ3)	Yes
	Barriers to delivery do not impede the programme being delivered with fidelity (RQ4)	Somewhat—challenge can be addressed. By being explicit about the expectations around cascading in the MoU and communicating these to senior leaders, the delivery team will increase the likelihood of schools putting aside the required time for cascading. By introducing a tiered approach, teachers will be able to adapt weekly problems and cater to pupils' different ability levels.
Evidence of promise	Teachers are motivated to engage in the programme and change classroom practice (RQ6)	Somewhat—challenge can be addressed. The challenge can be addressed by introducing tiered approaches to problems (see above) and by considering the suitability of each question for a debrief.
	The train-the-trainer approach results in lead trainers feeling confident in supporting the Year 7 teaching team (RQ7) The train-the-trainer approach results in practitioners feeling knowledgeable about key elements of the programme (RQ7)	Somewhat—challenge can be addressed but will require careful and detailed planning. This can be addressed by being more explicit about the requirements of cascading; spending more time on how to cascade effectively during the initial training session; introducing new resources for cascading (for example, workbooks); and by including more video exemplifications as part of cascading sessions, alongside stipulating the need for teachers to rehearse key programme activities such as the 'think aloud' approach.

	Teachers are more knowledgeable and confident in teaching tackling of unfamiliar mathematical problems (RQ8)	Somewhat—challenge can be addressed. This can start being addressed by tweaking the messaging about TM as a programme benefitting pupils <i>and</i> teachers. Through placing more emphasis on communication and training materials on the TM programme as a CPD programme, the delivery team can contribute to an awareness among programme participants that reflection about their own professional practice is essential for behaviour change.
	Pupils are motivated to engage in the programme (RQ9)	Somewhat—challenge can be addressed. This can be addressed by making problems more accessible through a ‘tiered’ approach, by considering the sequence of questions and considering the suitability of each question for a debrief.
	Positive change in pupils’ approaches to tackling unfamiliar mathematical problems (RQ10) No evidence that disadvantaged pupils benefit less than their less disadvantaged peers (RQ10)	Somewhat—challenge can be addressed. For the data to be more robust, it would be advisable to administer short surveys with pupils to complement teacher insights on pupils’ progress. Another option is to include a longitudinal design and observe the same class over several timepoints to better capture any changes in pupils’ approaches to tackling unfamiliar mathematical problems.
	Changed practice by school and teachers embedded throughout 20 weeks and intended to be sustained after 20 weeks (RQ11)	Yes
Readiness for trial	Core components identified for key programme activities (RQ12)	Yes
	Sufficient data collected to monitor compliance and fidelity (RQ13) Time spent on data entry (for example, completing delivery logs) is proportionate to its purpose (RQ13)	Somewhat—challenge can be addressed but will require careful and detailed planning. This can be addressed by making some changes to the way the delivery team captures data on in-school fidelity via lead teachers, enabling a better understanding of schools’ adherence to core components.
	Outcome measure(s) identified that is/are a) aligned to intended outcomes; b) able to measure changes that we expect to see; and c) convenient to administer in secondary schools (RQ14)	Yes
	Clear and realistic plan for what scaling to deliver in more schools would look like and require (RQ15) Plan for scaling would entail little or no modification to content of in-school delivery (RQ15)	Somewhat—challenge can be addressed but will require careful and detailed planning. It will be important for the delivery team to consider how best to make changes to the cascading model. This might include being more explicit about the requirements of cascading, introducing new resources for cascading (for example, workbooks), and including more video exemplifications as part of cascading sessions.

Conclusion

Table 15: Summary of pilot findings

Area of research	Findings
<p>Feasibility</p> <p>Is the approach feasible to implement?</p>	<p>The programme was partly delivered as intended. While training took place as planned and cascading took place in most schools, the programme was not consistently delivered as intended, with variation regarding session duration and coverage. Classroom sessions were delivered regularly, although teachers did not consistently adhere to the core components, with inconsistent usage of programme materials and strategies. Teachers praised the classroom resources for their clarity but highlighted that some questions were less suitable to pupils with special educational needs and disabilities (SEND) and English as an additional language (EAL). To address challenges to fidelity caused by inconsistent cascading, the delivery team has considered tightening up the cascading model, including communicating the requirements of cascading more clearly to senior leaders. To ensure mathematical problems are suitable for different pupil groups in the future, the delivery team are considering including differentiation, providing different difficulty levels for the same questions.</p>
<p>Evidence of promise</p> <p>Is there evidence to support the theory of change (ToC)?</p>	<p>Findings indicate that the programme has some evidence of promise, but that some causal pathways are weaker than others. Although the response rate to the post-delivery surveys was low, most respondents reported improved problem-solving for pupils as a positive outcome of the programme. Pupils and teachers perceived problem-solving to be different to ‘business as usual’ teaching, and pupils expressed being motivated to engage with the programme. However, teachers reported that pupil engagement was not consistently high and was shaped by how accessible and relevant pupils found the problems. Teachers were motivated to engage with the programme, although their engagement was negatively affected where pupils were disengaged. Pre-post analysis shows positive shifts in teachers’ understanding of and confidence in teaching problem-solving. However, when asked specifically about benefits resulting from the programme, only a small majority of teachers reported clear personal gains. Those teachers who identified improved confidence in interviews noted that the routine of teaching problem-solving following the same structure throughout the programme helped develop confidence. Post-delivery survey data indicate promising levels of sustainability. Nearly three-quarters of teachers reported intending to sustain key practices, including encouraging pupils to discuss and compare approaches (the debrief) and modelling their thinking through ‘think aloud’. However, it is important to note that the self-selecting nature of the respondents means these survey results may not represent the views of all pilot participants.</p>
<p>Readiness for trial</p> <p>Is the approach ready to be evaluated in a trial?</p>	<p>The training was not consistently and effectively cascaded to Year 7 teachers by lead teachers; given the importance of cascading in ensuring implementation fidelity, the intervention is not yet fully ready to be evaluated for a trial and would instead benefit from a further pilot evaluation. The programme is partially manualised: comprehensive and high-quality materials exist but require additional guidance and clearer articulation of core components before delivering the programme to a larger number of schools, including a more explicit distinction between non-negotiable and flexible elements. To ensure fidelity at a larger scale, the delivery team would benefit from a more refined system for monitoring all core components.</p> <p>The evaluation team identified the Progress Test in Maths (PTM) as an appropriate outcome measure for an efficacy trial. The problem-solving subscale raw score can be used as a secondary, exploratory outcome, as it aligns with the programme’s focus on problem-solving. PTM is UK normed and offers practical whole-class administration with centralised scoring.</p>

Formative findings

Across the interviews, lead teachers, teachers, pupils, as well as the delivery team, made several recommendations for refinements and improvements related to training and ongoing support and classroom delivery. Combined with our own reflections, these are presented below.

Training and ongoing support

Recommendation 1: Consider tweaking the messaging to position TM as a programme benefitting pupils *and* teachers.

Rationale and detail: By placing more emphasis on communication and training materials on TM as a CPD programme, the delivery team can contribute to an awareness among programme participants that reflection about their own professional practice is essential for behaviour change.

Recommendation 2: Consider providing schools more opportunities for peer-to-peer learning.

Rationale and detail: Opportunities for schools to share good practice and barriers can engender ‘buy-in’ and give schools practical solutions for addressing any challenges they might experience. Opportunities for sharing their implementation journey will be particularly beneficial for schools operating in similar contexts.

Recommendation 3: Consider ensuring a broadly equal balance in group size across in-person training sessions.

Rationale and detail: Lead teachers will likely benefit from conversations and informal discussions with a wide range of other schools, including schools with shared contextual challenges. Bigger training groups might also more likely engender a regional ‘community of practice’, enabling lead teachers to share insights with other lead teachers in their area throughout the programme. For the delivery team, equally balanced group sizes across regions will facilitate more consistent delivery and encourage more peer-to-peer learning during the sessions.

Weekly problems

Recommendation 4: Consider providing more guidance on the duration of each session.

Rationale and detail: Providing more guidance to schools including a breakdown of how much time should broadly be spent on each activity during a 20-minute session could ensure a) more consistent delivery across schools, and b) a reduction in impact on other elements of the maths curriculum if teachers don’t spend more time than required on TM sessions.

Recommendation 5: Consider making problems more accessible.

Rationale and detail: Weekly problems can be adapted and include different difficulty levels. This ‘tiered’ approach would enable more consistent engagement from high-ability and low-ability pupils while ensuring that pupils still tackle an unfamiliar mathematical problem. It would also ensure that problems can be addressed by schools with different maths curricula. To ensure problems are accessible for pupils with SEND and EAL, the delivery team may also want to consider incorporating visuals into the problem layout. The inclusion of tiered questions will likely require the training to focus more explicitly on how to address differentiation and how to ensure the debrief can work effectively where there are different questions for pupils to choose from.

Recommendation 6: Consider the sequence of questions.

Rationale and detail: By selecting easier questions at the beginning of delivery, teachers can more easily become accustomed to focusing on the process of problem-solving rather than spending time discussing the answer.

Recommendation 7: Consider the suitability of each question for a debrief.

Rationale and detail: The importance of the debrief for improving pupils’ mathematical problem-solving requires pupils to tackle a problem that offers different methods of problem-solving and generates a discussion between pupils.

'Think aloud', prompts, and debrief

Recommendation 8: Consider providing more explicit guidance about the use of prompts.

Rationale and detail: By ensuring that prompts are readily available (for example, by printing out the prompts or attaching them to the wall), teachers can ensure that all pupils have the tools to approach a problem systematically. It is also likely to help those pupils who might struggle with 'starting' a task.

Recommendation 9: Consider adopting the gradual release of responsibility model for the 'think aloud' approach.

Rationale and detail: By omitting the 'think aloud' approach from weekly sessions towards the end of programme delivery, teachers can put the 'onus' on pupils to take more responsibility for independent problem-solving while still supporting those pupils who 'get stuck'.

Recommendation 10: Consider providing more guidance on the debrief.

Rationale and detail: By providing clearer expectations of what a debrief looks like without being prescriptive about the debrief following one specific format, the delivery team are more likely to ensure that debriefs happen in every lesson. The explicit guidance will also likely benefit those teachers unfamiliar or less confident with this approach.

Outcome measure

Recommendation 11: Use the full PTM total score as the primary outcome and the problem-solving subscale score as a secondary, exploratory outcome.

Rationale and detail: The total score provides the most robust, validated indicator of overall attainment, while the subscale offers problem-solving insight. However, the subscales are not psychometrically designed to stand alone as outcome measures, hence its use as exploratory.

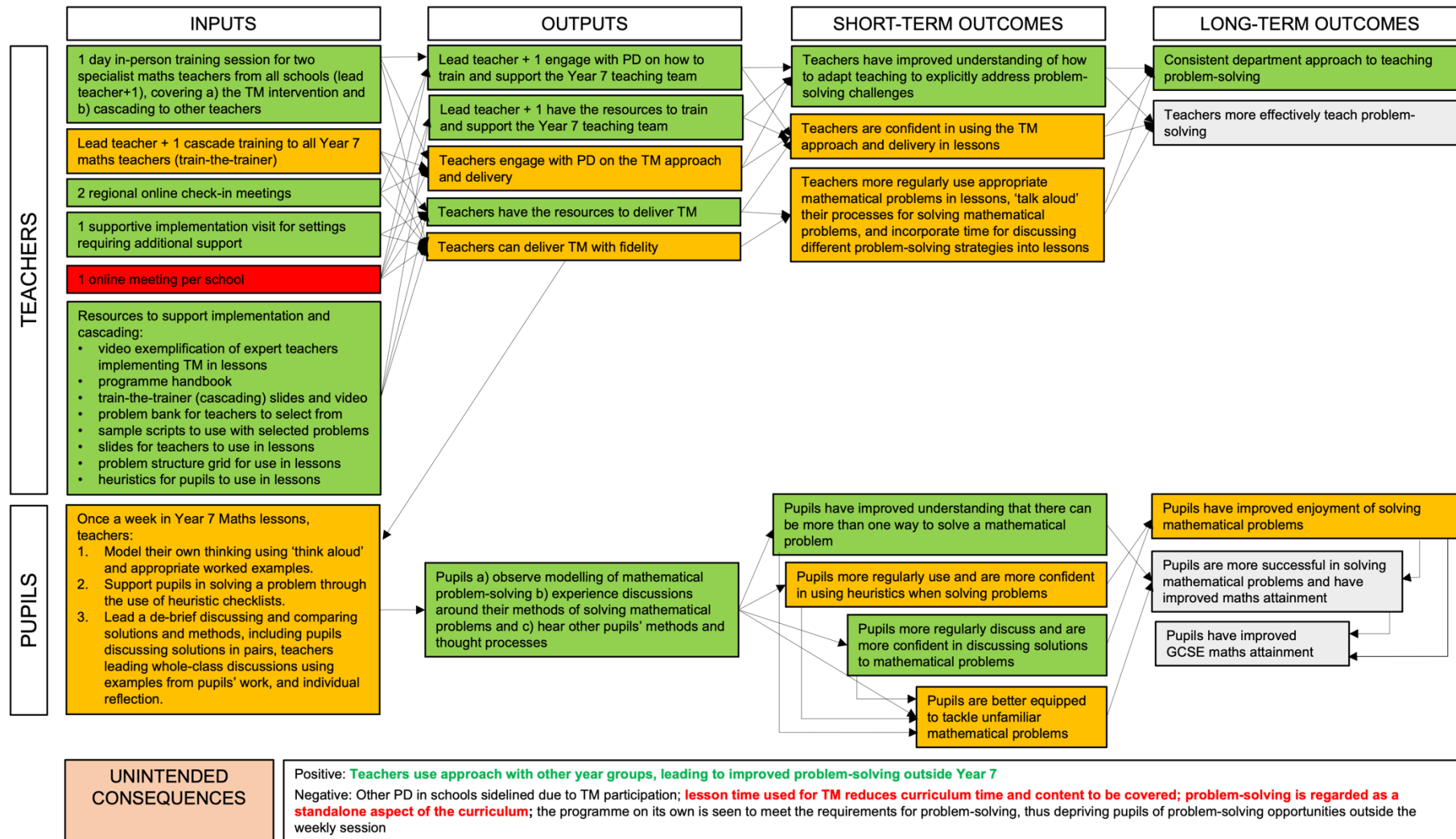
Recommendation 12: Use raw scores for the PTM rather than age-standardised scores as the outcome measure.

Rationale and detail: Raw scores avoid distortions introduced by age-norming and therefore offer greater sensitivity, especially at the lower end of the distribution, and help mitigate observed floor-effect risks. Hodgen *et al.* (2019) report that apparent floor effects did not occur in the raw PTM data but emerged during age standardisation, indicating that norming compressed lower-attaining pupils' scores. Using raw scores therefore preserves real variation, reduces artificial clustering at zero or very low scores, and provides a truer picture of change.

Interpretation

Figure 13 presents a visual overview of the strength of evidence for different components of the ToC, showing a mixed picture with regard to implementation fidelity and evidence of promise. The following sections on feasibility and evidence of promise provide a detailed explanation for the decision to colour-code different elements of the ToC green, yellow, and red.

Figure 13:³² Strength of evidence



³² For **inputs** and **outputs**, elements highlighted in green indicate high fidelity, yellow indicates partial fidelity, and red signifies limited fidelity. For **outcomes**, elements highlighted in green indicate promising evidence of promise while yellow suggests the evidence is more mixed. Boxes in grey indicate that either no or insufficient evidence was collected to make an informed judgement. For **unintended consequences**, text in bold highlights some evidence in support of the manifestation of a positive and/or negative consequence.

Feasibility of implementation

Findings suggest that parts of the intervention were feasible for participants to implement, with challenges remaining around the model of cascading learning to Year 7 teachers and the suitability of mathematical problems.

In-person training was well attended, highlighting senior leaders' willingness to release teachers for training. Although attendance of **regional check-in meetings** was patchier than the in-person training, it appears that many schools did not deem it necessary for both lead teachers to be present. While there is limited detail from schools directly about the **implementation support** visits, there is no indication from our findings that they cannot be accommodated by schools. The delivery team was conscious that the in-person mode of implementation support visits meant they needed to be aimed at those schools that needed the most support.

The whole-cohort delivery model makes a **cascading** model necessary for the TM programme. The programme would not be feasible without a cascading model, as it would be impractical for *all* participants to attend in-person training. However, in its current form, the cascading model risks that the programme is not delivered as intended across schools. It is very likely that the inconsistent implementation of classroom sessions across schools, reflected in participants' inconsistent usage of programme materials or the debrief, is related to the different ways lead teachers cascaded learning with their Year 7 teachers. Discussions with the delivery team during the learning workshop demonstrated the need for more training on how to effectively cascade as well as tightening the cascading requirements. This can mitigate the risks inherent in cascading models, including 'dilution' of quality, the omission of key information, and teachers receiving the 'cascaded learning' who might be less confident and experienced than those attending the training (see also Hayes, 2000). Changes to the cascading model and guidance will likely result in classroom sessions being delivered with more fidelity, including closer adherence to the session duration requirement and more frequent use of the **debrief**. However, changes to the cascading model and guidance can only go so far; it is equally important for SLT to protect time so that cascading sessions can be delivered as required.

The **content and the implementation approach** were largely acceptable to participants. Participants praised the resources for their relevance and clarity. This is particularly important because maths in secondary schools can be taught by non-subject specialists (Boylan *et al.*, 2024; see also Sani and Burghes, 2021). Ready-made scripts in particular helped teachers implement the classroom sessions with minimal preparation required. However, participants were more critical of the selected problems, as they did not consider all to be appropriate for all pupils in their class, notably pupils with SEN or EAL or for high-attaining pupils. Some schools even discontinued the programme for their lowest attaining ('bottom set') pupils. Because the programme is targeted at *all* Year 7 classes, this finding is perhaps less surprising; nonetheless, it points to the importance of considering more carefully how to differentiate so that all pupils can access the problems. The delivery team's reflections, such as including a 'tiered' approach to presenting weekly problems, demonstrate that they have acknowledged the need to tweak this programme element.

Evidence of promise

While the findings demonstrate some promising evidence to support the theory of change—in particular around perceived changes in pupil behaviour and practice—the evidence is more varied for other elements of the ToC.

The ToC posits that teachers have **improved understanding of how to teach problem-solving**. The quantitative findings point to promising evidence in this area; in the post-delivery survey, slightly more than half of survey respondents reported improved understanding of how to teach problem-solving as a positive outcome of the programme, and pre-post analysis of questions confirmed improvements in lead teachers' and Year 7 teachers' understanding in teaching problem-solving. This finding relates directly to a key gap identified in the ToC, which is that most secondary maths departments do not have a **consistent way of supporting pupils to develop as problem-solvers** across their team of teachers (see also Boylan *et al.*, 2024). This was reflected in our pre-delivery survey, where only one in five lead teachers and Year 7 teachers reported that approaches to problem-solving were 'consistent' or 'very consistent' across teachers at their school, underlining the need for a more consistent approach to problem-solving across pilot schools. Post-delivery survey findings suggest that **there is a more consistent departmental approach to teaching problem-solving**, with nearly two-thirds of participants who initially reported having 'not at all' or 'only slightly' consistent approaches to teaching problem-solving at their school reporting 'consistent' or 'very consistent' approaches.

Teachers broadly felt **confident in using the TM approach and delivery in lessons**. Pre-post analysis of questions confirmed improvements in lead teachers' and Year 7 teachers' confidence in teaching problem-solving and confidence in facilitating discussions. In the post-delivery survey, two in three survey respondents reported increased confidence in teaching problem-solving as a positive consequence of programme participation. And in qualitative interviews, lead teachers and Year 7 teachers attributed the improved confidence in teaching problem-solving to the resources, which reminded them to model their thinking and focus on the process of arriving at an answer rather than the answer itself. This supports a key causal pathway in the ToC, which is that teachers having access to TM resources leads to improved confidence in using the TM approach and delivery in lessons. The absence of suitable resources to teach problem-solving, identified as a challenge by lead teachers and Year 7 teachers in the pre-delivery survey, likely makes it difficult for teachers to be confident in teaching it.

In the pre-post analysis of questions, participants reported a **slight decline in confidence in modelling their own thinking and confidence in identifying appropriate problems**. We suggest that it is possible that these negative shifts reflect teachers realising previously unrecognised gaps in their practice, correcting an initial overestimation once they better understood 'what they didn't know' through participation in the programme. This explanation is particularly plausible when we consider wider evidence suggesting that problem-solving is felt to be a particular area of weakness in the curriculum (Noyes *et al.*, 2023), with secondary maths leaders highlighting problem-solving as a key professional development priority (Boylan *et al.*, 2024).

The evidence that teachers more regularly use appropriate mathematical problems in lessons, 'talk aloud' their processes for solving mathematical problems, and incorporate time for discussing different problem-solving strategies into lessons is patchy. Part of the reason why core components—notably the debrief—have not systematically been adhered to across pilot schools relates to some schools not cascading effectively. As a result, they missed an opportunity to adopt behaviours that drive effective implementation; as highlighted in the EEF guidance on effective implementation (EEF, 2021), this includes uniting people around what is being implemented, how it will be implemented, and why it matters.

There were perceived improvements from pre- to post-delivery in pupils' understanding that there can be **multiple solutions to a problem**, in **pupils' confidence to solve mathematical problems**, in **pupils' confidence discussing problem-solving**, and in **how pupils reflect on how to approach a problem**. Further, more than two-thirds of post-delivery survey respondents explicitly identified 'improved problem-solving' in pupils as a positive outcome of the programme. These findings suggest promise in the programme changing pupil practice and demonstrate that providing regular opportunities for pupils to explore strategies for problem-solving can enhance their confidence and resilience (Ofsted, 2021). However, interviews with lead teachers and Year 7 teachers highlighted challenges around the accessibility of problems for pupils who were less secure in maths or EAL pupils; this suggests that the delivery team will need to think carefully how to ensure that the mathematical questions cater to pupils' varying proficiency levels in maths.

A recurring theme in the qualitative research with pupils was the perspective that the programme was different to 'business as usual' because Year 7 teachers delivering the sessions emphasised the process of arriving at an answer and encouraged different approaches rather than focusing on the actual answer itself (see also Liljedahl *et al.*, 2016). This adds to the evidence that pupils in UK secondary schools have limited exposure to practising solving problem-solving questions in class (Boylan *et al.*, 2024). Our findings show that pupils appreciated that TM sessions were about nuanced discussions which encouraged mistakes, something that teachers saw reflected in pupils developing a problem-solving mindset.

A more medium-term outcome in the ToC is pupils developing **improved enjoyment of problem-solving** and becoming **better at solving unfamiliar mathematical problems**. Pre-post analysis did not find improvements for these two perceived outcomes, suggesting that they take longer to materialise. It is possible that a wider range of mathematical problems suited to different pupils' needs as well as a longer implementation delivery window would make it more likely for these perceived outcomes to materialise.

Readiness for trial

Challenges around the cascading model and, relatedly, the inconsistent adherence to core components suggest that the intervention is not fully ready to be evaluated for a trial and would instead benefit from further development and piloting.

The number of schools the delivery team will need to reach for an efficacy trial (minimum of 50 schools) is significantly bigger than the number of schools involved in the pilot (30 schools), but the delivery team are confident they can recruit and deliver training at this scale. Their plans appear realistic, although not refined at this stage.

For classroom delivery, the **core components** that are essential for the intervention's success, if implemented successfully, are clear. However, what is currently missing from the programme resources is clarity and what types of adaptations are permitted and encouraged, which is essential given the importance that contextual factors can play in shaping delivery (EEF, 2021).

Findings suggest that the **monitoring systems** the delivery team used to monitor fidelity did not sufficiently capture all criteria for fidelity. This would be particularly essential when delivering at a larger scale, and so the delivery team would benefit from refining their systems for collecting data from schools so that they can effectively support implementation (see Hall *et al.*, 2025).

Because Year 7 teachers seemed less prepared about delivery following cascading than lead teachers, it will be important for the delivery team to consider how best to make changes to the **cascading model**. This might include being more explicit about the requirements of cascading, introducing new resources for cascading (for example, workbooks), and including more video exemplifications as part of cascading sessions. Because the success of the TM programme is largely dependent on Year 7 teachers understanding the programme and delivering the programme as intended, we believe it would be too 'risky' to proceed to trial without being confident in the robustness of an adapted cascading model.

Limitations

There are several limitations within the study, which need to be considered when interpreting the findings.

The evaluation was a pilot evaluation, and so quantifying impact at either teacher or pupil level was beyond the scope of the study. While we can conclude that the findings indicate that the programme might have an impact if it goes to trial, we cannot state this is the case.

During case study visits, there were instances where Year 7 teachers believed that the evaluation team was working with the delivery team. For instance, it is possible that they could have been more critical of the programme if they understood prior to the interview that the evaluator was distinct from the delivery team. As an evaluation team, we will reflect on how we can better communicate to participants at different stages of the evaluation that we are distinct from the delivery team.

We had limited control over which children the lead teachers selected for the pupil focus groups or which of the sessions we observed on the day. It is likely that teachers 'cherry-picked' pupils who were more engaged in the programme than their peers. Moreover, we did not carry out other data collection activities with pupils, such as short surveys, which could have strengthened some of the findings from pupil interviews and particularly complemented teachers' perceptions on pupils' outcomes.

Similar to pupils being 'cherry-picked' by teachers, it is possible that lead teachers wanted us to observe teachers delivering sessions who were more confident and knowledgeable about problem-solving than the 'typical' maths teacher at their school. While this suggests that observation data should be treated with caution, it is important to be mindful that evaluators also need to be pragmatic when carrying out research in schools.

The respondent profile of lead teachers and Year 7 teachers for the pre- and post-delivery survey had a high proportion of subject specialists; it is possible that a higher proportion of non-subject specialists would have yielded different responses, particularly around understanding and confidence to teach problem-solving. Similarly, because fewer participants

completed the post-delivery survey, the pre-post comparisons may not fully reflect non-respondents' views and should therefore be interpreted cautiously.

The deadline for completing the post-delivery survey was extended due to an unsatisfactory response rate. This meant that there was a two-month time gap between some participants completing the post-delivery survey and completing the programme; it is possible that participants' recollections about some parts of the programme were affected by this time gap.

Future research and publications

We suggest that a future trial uses the PTM total score as the primary outcome and the problem-solving subscale score as a secondary, exploratory outcome. The total score provides the most robust, validated indicator of overall attainment, while the subscale offers problem-solving insight. However, the subscales are not psychometrically designed to stand alone as outcome measures, hence its use as exploratory.

A future pilot evaluation or trial should consider exploring how cascading happens in more depth—including whether any tightened requirements are met. This might include carrying out short interviews with lead teachers and Year 7 teachers immediately after their cascading sessions to understand what happened, how, and why; it might also include the observation of cascading sessions.

We also recommend interviews with senior leaders beyond interviews with lead teachers. This would recognise the crucial role that senior leadership plays alongside lead teachers in ensuring effective implementation of TM, specifically by adopting the behaviours that drive effective implementation, attending to the contextual factors that influence implementation, and using a structured but flexible implementation process.

There are no planned publications beyond the evaluation report.

References

- Action for Children (2024) 'Where is Child Poverty Increasing in the UK?' Action for Children. Available at: <https://www.actionforchildren.org.uk/blog/where-is-child-poverty-increasing-in-the-uk/> (accessed 22 July 2025)
- Basharat, M., Taylor, I., Duysak, E., and Kuo T. L. (2023) Stop and Think: Learning Counterintuitive Concepts: Statistical Analysis Plan. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/EEF-SAP-Stop-and-Think_v1.pdf?v=1723586444 (accessed 28 August 2025)
- Boylan, M., Zhu, H., Jaques, L., Birkhead, A., and Rempe-Gillen, E. (2024) Secondary Maths Practice Review. London: Education Endowment Foundation. Available at: <https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/secondary-maths-practice-review> (accessed 22 July 2025)
- Brady, L. M., and Graham, B. (2019) *Social research with children and young people*. Bristol: Policy Press.
- Business in the Community (BITC) (2023a) Regional Insights on Race Factsheet: North West. London: Business in The Community. Available at: <https://www.bitc.org.uk/wp-content/uploads/2023/03/bitc-factsheet-race-raw-regional-insight-north-west-march23.pdf> (accessed 22 July 2025)
- Business in the Community (BITC) (2023b) Regional Insights on Race Factsheet: Yorkshire and the Humber. London: Business in The Community. Available at: <https://www.bitc.org.uk/wp-content/uploads/2023/03/bitc-factsheet-race-raw-regional-insight-yorkshire-humber-march23.pdf> (accessed 22 July 2025)
- Crenna-Jennings, W. (2018) *Key drivers of the disadvantage gap: Literature review*. London: Education Policy Institute. Available at: <https://epi.org.uk/wp-content/uploads/2018/07/EPI-Annual-Report-2018-Lit-review.pdf> (accessed 20 May 2026)
- Curriculum and Assessment Review (2025) Building a world-class curriculum for all: Final report. November 2025. Available at: https://assets.publishing.service.gov.uk/media/690b96bbc22e4ed8b051854d/Curriculum_and_Assessment_Review_final_report_-_Building_a_world-class_curriculum_for_all.pdf (accessed 09 December 2025)
- Demack, S., Culliney, M., Boylan, M., and Wolstenholme, C. (2022) Realistic Maths Education: Evaluation report. Education Endowment Foundation. Available at: <https://shura.shu.ac.uk/31112/1/Realistic-Maths-Education-FINAL.pdf> (accessed 16 October 2025)
- Department for Education (DfE) (2021) 'National curriculum in England: mathematics programmes of study'. 28 September 2021. Available at: <https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study/national-curriculum-in-england-mathematics-programmes-of-study> (accessed 09 December 2025)
- Department for Education (DfE) (2023a) Guidance: Education Investment Areas. GOV.UK. Available at: <https://www.gov.uk/government/publications/education-investment-areas/education-investment-areas> (accessed 07 June 2025)
- Department for Education (DfE) (2023b) 'What are academy schools and what is 'forced academisation''? 02 May 2023. Available at: <https://educationhub.blog.gov.uk/2023/05/what-are-academy-schools-and-what-is-forced-academisation/> (accessed 01 December 2025)
- Department for Education (DfE) (2025) Academic Year 2024/25: Schools, Pupils, and Their Characteristics. GOV.UK. Available at: <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics/2024-25#dataBlock-d49543e4-9c0f-4456-bb2f-0f3383777c83-tables> (accessed 22 July 2025)
- Education Policy Institute (EPI) (2024) Ofsted Ratings—Have Standards Been Improving? Accountability, Assessment, and Inspection Blog. EPI. Available at: <https://epi.org.uk/publications-and-research/ofsted-ratings-have-standards-been-improving/> (accessed 20 May 2026)

- EEF (2018) Metacognition and Self-regulated Learning. Available at:
https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/metacognition/EEF_Metacognition_and_self-regulated_learning.pdf (accessed 09 September 2025)
- EEF (2021) Effective Professional Development: Guidance Report. Available at:
<https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/effective-professional-development/EEF-Effective-Professional-Development-Guidance-Report.pdf> (accessed 15 October 2025)
- EEF (2022a) Implementation and Process Evaluation Guidance for EEF Evaluations. Available at:
<https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/EEF-IPE-Guidance-August-2022.pdf?v=1660741017> (accessed 08 August 2024)
- EEF (2022b) Improving Mathematics in Key Stages 2 and 3: Guidance Report. Available at:
<https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/maths-ks-2-3/EEF-Improving-Mathematics-in-Key-Stages-2-and-3-2022-Update.pdf> (accessed 17 October 2025)
- EEF (2023) Guidance for EEF Pilot Evaluations. Available at:
<https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Pilot-guidance-Oct-2023.pdf?v=1721031784> (accessed 22 January 2025)
- EEF (2024) A School's Guide to Implementation: Guidance Report. Available at:
https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/implementation/a_schools_guide_to_implementation.pdf (accessed 05 December 2025)
- EEF (n.d.) 'Integrating evidence into maths teaching: Planning a Think Aloud'. Available at:
https://d2tic4wvo1iusb.cloudfront.net/eef-guidance-reports/maths-ks-2-3/Planning_a_Think_Aloud_1.0.pdf (accessed 19 August 2025)
- GL Assessment (n.d.) Mathematics process categories: Form A. Accessible at: https://support.gl-assessment.co.uk/media/2980/ptm7-links-to-national-curricula_form-a.pdf (accessed 31 August 2025)
- Groom, M., Nightingale, M., Flemons, L., Merewood, J., Amodeo, G., Oades, F., Senator, B., and Speciani, E. R. (2024) Improving Working Memory plus Arithmetic: Evaluation report. Education Endowment Foundation. Available at:
<https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/IWMMA-final-report.pdf?v=1760407017> (accessed 16 October 2025)
- Hall, A., Bean, A. M., Taylor, P., and Lewis, J. (2025) EEF Scaling Framework. Available at:
https://d2tic4wvo1iusb.cloudfront.net/production/documents/EEF_scaling-framework_v.1.1.0.pdf?v=1760017512 (accessed 09 October 2025)
- Hayes, D. (2000) 'Cascade training and teachers' professional development'. *ELT Journal*, 54(2), 135–145.
- Hiebert, J. and Grouws, D. A. (2007) 'The Effects of Classroom Mathematics Teaching on Students' Learning'. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 371–404). Charlotte, NC: Information Age.
- Hodgen, J., Taylor, B., Anders, J., Tereshchenko, A., and Francis, B. (2019) The Student Grouping Study: Investigating the effects of setting and mixed attainment grouping. Study Plan. Education Endowment Foundation. Available at:
https://discovery.ucl.ac.uk/id/eprint/10117537/1/Students_Grouping_Study_Study_Plan_2019.11.19.pdf (accessed 20 May 2026)

- Homes England (2023) Fact Sheet 5: New Homes and School Places. GOV.UK. Available at: <https://www.gov.uk/government/publications/new-homes-fact-sheet-5-new-homes-and-school-places/fact-sheet-5-new-homes-and-school-places> (accessed 22 July 2025)
- Humphrey, H., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., and Kerr, K. (2016) *Implementation and process evaluation (IPE) for interventions in educational settings: An introductory handbook*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/IPE_Handbook.pdf?v=1630567508 (accessed 20 May 2026)
- Jay, T., Thomas, P., and Demack, S. (2016) Dialogic Teaching: Improving classroom talk. Statistical Analysis Plan. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Round_6_-_Dialogic_Teaching_SAP.pdf?v=1746743360 (accessed 28 August 2025)
- Joseph Rowntree Foundation (JRF) (2025) UK Poverty 2025. York: Joseph Rowntree Foundation. Available at: <https://www.jrf.org.uk/uk-poverty-2025-the-essential-guide-to-understanding-poverty-in-the-uk#:~:text=Figure%203:%20Poverty%20rates%20vary%20significantly%20between%20UK%20nations%20and%20regionsandtext=In%20the%20latest%20data%2C%20the,renters%20pay%20over%2080%25%20more> (accessed 22 July 2025)
- Leavy, A. M. and Hourigan, M. (2020) 'Expanding the focus of early years mathematics education: statistics and probability'. In Saracho and Spodek (Eds.) *Handbook of Research on the Education of Young Children* (pp. 99–112). London: Routledge.
- Lester Jr., F. K. (2013) 'Thoughts about research on mathematical problem-solving instruction'. *The Mathematics Enthusiast*, 10(1–2), 245–278.
- Liljedahl, P., Santos-Trigo, M., Malaspina, U., and Bruder, R. (2016) Problem solving in mathematics education. Springer Open. Available at: <https://library.oapen.org/bitstream/handle/20.500.12657/27743/1/1002262.pdf> (accessed 09 October 2025)
- Marshall, L., Bury, J., and Bashir Malik S. (2024) Thinking Mathematically: Study Plan. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/thinking_mathematically_-_study-plan.pdf?v=1753347509 (accessed 24 July 2025)
- Maths Horizons (2025) How England Should Reform Maths Education for the Age of AI. London: Maths Horizons. Available at: <https://www.mathshorizons.uk/reports> (accessed 15 May 2025)
- Mercer, N. and Sams, C. (2006) Teaching children how to use language to solve maths problems. *Language and education*, 20(6), 507–528.
- Noyes, A., Brignell, C., Jacques, L., Powell, J., and Adkins, M. (2023) The Mathematics Pipeline in England: Patterns, Interventions, and Excellence. Nottingham: Centre for Research in Mathematics Education (CRME). Available at: <https://nottingham-repository.worktribe.com/output/21372570/the-mathematics-pipeline-in-england-patterns-interventions-and-excellence> (accessed 17 January 2024)
- Office for Health Improvement & Disparities (2021) Health Profile for Yorkshire and the Humber 2021. London: Office for Health Improvement & Disparities. Available at: https://fingertips.phe.org.uk/static-reports/health-profile-for-england/regional-profile-yorkshire_and_the_humber.html (accessed 22 July 2025)
- Office for National Statistics (ONS) (2021) What are the Regional Differences in Income and Productivity? Office for National Statistics. Available at: <https://www.ons.gov.uk/visualisations/dvc1370/> (accessed 22 July 2025)
- Ofsted (2021) Research review series: mathematics. Available at: <https://www.gov.uk/government/publications/research-review-series-mathematics> (accessed 11 October 2025)

- Ofsted (2023) Coordinating mathematical success: the mathematics subject report. Available at: <https://www.gov.uk/government/publications/subject-report-series-maths/coordinating-mathematical-success-the-mathematics-subject-report> (accessed 11 October 2025)
- Perry, T., Lea, R., Jørgensen, C. R., Cordingley, P., Shapiro, K., and Youdell, D. (2021) Cognitive science in the classroom. London: Education Endowment Foundation (EEF). Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/guidance/Cognitive_Science_in_the_classroom_-_Evidence_and_practice_review.pdf?v=1701059743 (accessed 19 August 2025)
- Ritchie, J., Lewis, J., Nicholls, C. M., and Ormston, R. (eds.) (2013) *Qualitative Research Practice*. London: Sage.
- Russell, S. J. (2000) 'Developing computational fluency with whole numbers'. *Teaching Children Mathematics*, 7(3), 154–154.
- Sani, N. and Burghes, D. (2021) 'Longitudinal study of 'retraining' non-maths specialist teachers to become capable, confident teachers of mathematics'. *International Journal of Mathematical Education in Science and Technology*, 53(9), 2438–2464.
- Schools Week (2022). 'Nearly half of secondary schools using non-specialists to teach maths'. 22 November 2022. Available at: <https://schoolsweek.co.uk/nearly-half-of-secondary-schools-using-non-specialists-to-teach-maths/> (accessed 28 August 2025)
- The Council of Skills Advisors (n.d.) Council of Skills Advisors' Report. Skills Builder Partnership. Available at: https://labour.org.uk/wp-content/uploads/2022/10/WR-16813_22-Labour-Skills-Council-report-Edit-19-10-22.pdf (accessed 16 October 2025)
- Unity Research School (2025). 'What are some of the drivers of the attainment gap in mathematics at the end of Key Stage 2 for disadvantaged children? Links between language and literacy and disadvantage.' 02 February 2025. Available at: <https://researchschool.org.uk/unity/news/what-are-some-of-the-drivers-of-the-attainment-gap-in-mathematics-at-the-end-of-key-stage-2-for-disadvantaged-children> (accessed 19 August 2025)
- Walsh, M. (2016) 'Pre-service primary teachers' understanding of mathematical problem posing and problem solving: exploring the impact of a study intervention'. Mary Immaculate College, University of Limerick. Available at: <https://dspace.mic.ul.ie/bitstream/handle/10395/2089/Walsh%2c%20Maura%20%282016%29%20Pre-service%20Primary%20teachers%20understanding%20of%20mathematical%20problem%20posing%20and%20problem%20solving.%20MA.pdf?sequence=3&isAllowed=y> (accessed 09 October 2025)
- Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R., and Ogbuehi, P. (2012) *Improving Mathematical Problem Solving in Grades 4 through 8*. IES Practice Guide. NCEE 2012-4055. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. What Works Clearinghouse.

Appendices

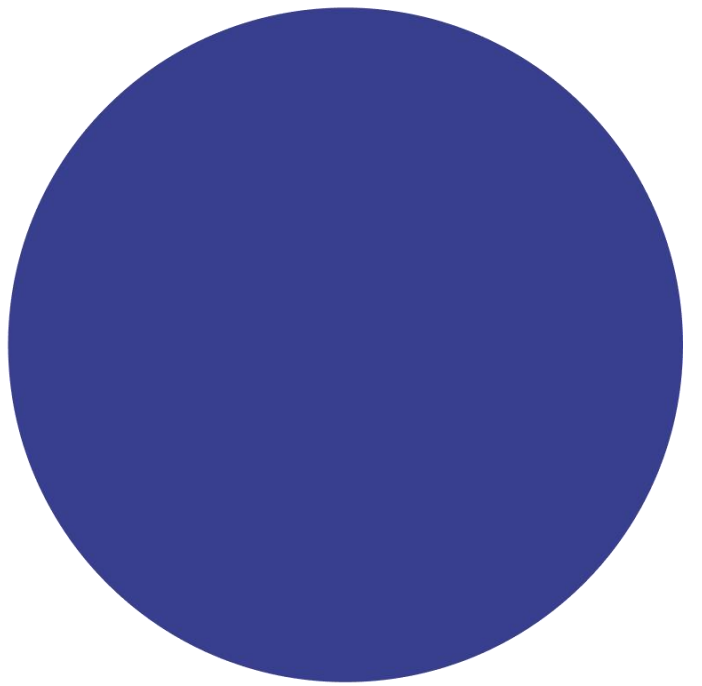
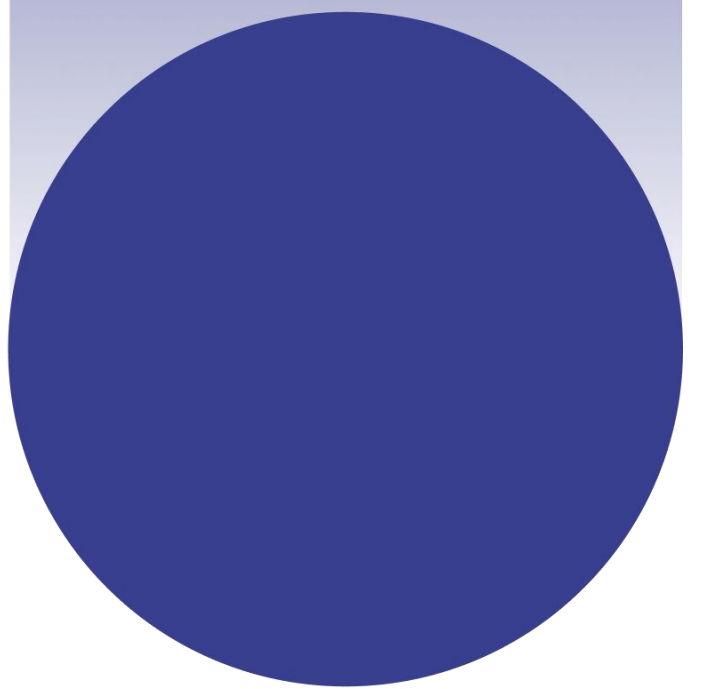
The appendices can be found published separately on the project page.

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit <https://nationalarchives.gov.uk/doc/open-government-licence/version/3> or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third-party copyright information, you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.


This document is available for download at <https://educationendowmentfoundation.org.uk>




Education
Endowment
Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

<https://educationendowmentfoundation.org.uk>

 @EducEndowFoundn

 Facebook.com/EducEndowFoundn