Maths Through Picture Books

Pilot report

July 2025

Jonah Bury, Sara Bashir Malik, Robert Wishart, and Lydia Marshall

The Education Endowment Foundation (EEF) 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
London
SW1P 4QP

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

This evaluation was conducted by a team from Oxford MeasurEd that included: 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); and Paulina Valenzuela (Junior Evaluator and Assessment Expert). The team was supported during fieldwork by Claire McAlpine (formerly Senior Consultant at Oxford MeasurEd) and Ross Duncan (Associate).

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 (Project Director)
Oxford MeasurEd
27 Mortimer Street
London
W1T 3BL

Email: Lydia.marshall@oxfordmeasured.co.uk

Executive summary

The project

Maths Through Picture Books (MTPB) is a targeted intervention to improve maths attainment among reception-aged children identified as being at risk of not developing a secure understanding of number before they move to Key Stage 1. It equips teachers and teaching assistants (TAs) with improved knowledge of how children learn number and operations. Moreover, teachers and TAs gain the knowledge and skills to engage a small group of targeted children in sustained multi-turn conversations about number and operations using picture books and the ShREC (Share attention, Respond, Expand, Conversation) approach. Through the programme, children are expected to improve their enjoyment and motivation towards learning maths and to talk about number and operations. A team from East London Research School (ELRS) delivers training to one teacher and one TA per school. Training involves a full-day in-person training session at the beginning of delivery and a half-day in-person training session at the end of delivery. Additionally, there is an online session for teachers on supporting their TAs, and an online drop-in session for all teachers and TAs to support delivery. While TAs and teachers are trained, TAs are responsible for direct delivery to children.

The MTPB intervention is delivered face-to-face to groups of five children aged four to five years, identified as requiring extra support in maths based on a short task and accompanying guidance. MTPB sessions are delivered twice per week for ten weeks. The sessions last for 20 minutes and occur at the same time every week in a quiet, dedicated space away from the classroom. TAs run interactive reading sessions with a set of nine specially chosen maths-focused picture books. Conversations are scaffolded using the ShREC approach to encourage back and forth interactions about the book. The TA and teacher have weekly planning meetings to ensure fidelity to the programme and consistency between the targeted sessions and wider classroom practice. During delivery, teachers observe the TA delivering the programme three times. The teachers then have a professional conversation with the TA on the same day, providing feedback and supporting the TA to reflect on the session using the ShREC framework. The delivery team designed a framework to break the strategies of the approach down into observable behaviours.

A total of 38 practitioners (19 teachers and 19 TAs) and 95 children from 19 schools took part in the pilot of the programme, which was delivered in schools between December 2023 (pupil selection) and June 2024 (final training day). One additional school withdrew following baseline testing in January 2024. The schools were based in the London boroughs of Newham, Havering, and Redbridge, as well as Essex. Evaluation activities included qualitative interviews with the delivery team, teachers, TAs, and senior leadership team (SLT) members, and two short online surveys with teachers and TAs. The evaluation team also carried out desk-based research, piloting of outcome measures, and analysis of programmatic data. Finally, the evaluation team observed training sessions and MTPB delivery.

Table 1: Summary of pilot findings

Research question	Finding
Feasibility Is the approach feasible to implement?	The programme was largely delivered with fidelity. Sessions took place in nearly all schools twice a week during the delivery period and nearly all TAs who completed the survey confirmed the sessions were delivered in a dedicated quiet space. Qualitative interviews with teachers and TAs highlighted that TAs nearly all adopted the ShREC approach, although observations showed that there was some variation in the extent to which they consistently followed it. Challenges to in-school fidelity are mainly related to the ongoing support provided by teachers for TAs. For instance, while planning and reflection sessions happened in nearly all schools, the coverage of the content and quality of support was inconsistent, with some teachers not probing sufficiently or giving little space to their TAs to share their reflections. To address this inconsistency in implementation, the delivery team is considering recruiting mentors to support in-school delivery in the future. Teachers and TAs surveyed were positive about the training and support from the delivery team, and this was mirrored in qualitative interviews with teachers and TAs who highlighted the practical relevance and novelty of the ShREC approach and praised the delivery team's delivery and approachability.
Evidence of promise Is there evidence to support the theory of change?	The quantitative and qualitative findings indicate that the programme has evidence of promise for practitioners and children. The overall majority of teachers and TAs taking part in the survey reported their own understanding of how children learn maths had improved. Similarly, most teachers and TAs who completed the survey reported their own confidence in teaching maths had improved. Qualitative interviews corroborated these

	findings, with teachers and TAs attributing the increased confidence in teaching maths to their exposure to a targeted and specific continuous professional development programme. The majority of teachers and TAs surveyed observed improvements in enjoyment and motivation towards learning maths in all participating children. In qualitative interviews, practitioners pointed to an increase in children's enjoyment and motivation as evidenced by high levels of engagement during the sessions. They identified the ShREC approach, play-based learning, and the small group size as key drivers.
Readiness for trial Is the approach ready to be evaluated in a trial?	The programme is ready to be evaluated in a trial. The delivery team and participants highlighted the importance and acceptability of the small group setting, the engaging medium of picture books for teaching numeracy, and the twice weekly 20-minute sessions. The delivery team have a clear and realistic plan for what scaling to deliver in more schools would look like and require, and have begun planning for this. Our desk-based research identified an appropriate outcome measure for a possible trial, and pilot data confirmed the measure's effectiveness in capturing the changes we expected to see, showing no floor or ceiling effects at baseline or endline. Feedback from teachers and assessors indicated that the Wechsler Individual Achievement Test-Third Edition UK (WIAT-III UK) Maths Problem Solving used at baseline and endline was generally acceptable and could therefore, be used in a trial. However, the Numeracy subtest was viewed as less suitable for use in schools due to its paper and pencil format and abstract number problems. We therefore, recommend that only the Maths Problem Solving subtest should be considered as an appropriate outcome measure for a trial.

Additional findings

The use of picture books and the ShREC approach appear to be sustained. Nearly all teachers and TAs who completed the post-delivery survey planned to continue using picture books to teach maths and to conduct small group reading sessions. The longitudinal follow-up survey showed that nearly all respondents reported to have sustained the use of picture books after ten weeks, although it is important to treat the findings with caution given the self-reported nature of the data and the small sample size, particularly for the longitudinal follow-up survey. All teachers who completed the survey found the guidance for selecting children for the intervention 'very useful', 'useful', or 'somewhat useful'. However, a notable minority (six out of 16 teachers who completed the survey) reported that only 'some' of the selected children were the right children. In qualitative interviews with teachers, TAs, and SLT members, participants who did not think the right children were selected attributed this to children being too 'able' or not all factors were being considered when selecting children. Qualitative interviews suggested that schools tended to include some disadvantaged children; however, the delivery team proposed that more schools could have included children with English as an Additional Language or children with Special Educational Needs. The evaluation did not find any evidence, which would indicate the programme would lead to a widening of the attainment gap between disadvantaged pupils and their peers.

Teachers and TAs perceived that children's socio-emotional learning improved in addition to their learning and communication development. In particular, the small group size and the emphasis on child-led communication were seen to contribute to improvements in children's language and communication. There were no major concerns about negative unintended consequences. The majority of teachers and TAs who completed the post-delivery survey believed that MTPB was different to the way maths was normally taught in reception. Teachers and TAs reflected that MTPB was more child-led than their usual approaches, creating space for children's opinions. Crucially, because MTPB discouraged practitioners from asking many questions, a group of TAs needed to unlearn their usual approach to communicating with children.

Introduction

Background evidence

In maths, the gap between low and high attainers in England is very wide by international standards (Ofsted, 2021). Gaps that are apparent by the end of early years double by the end of primary school (DfE, 2023a). Most children do not catch up by the end of compulsory education and this is particularly true for disadvantaged children (Hodgen *et al.*, 2018). Professional development programmes may help address this; however, in order for these programmes to be effective, they need to build educators' knowledge, motivate staff, develop teaching techniques, and embed this learning into everyday practice (Collin and Smith, 2021). For example, educators' knowledge can be built upon by revisiting prior learning, staff can be motivated by being presented information from a credible source, teaching techniques can be improved through modelling, and practice can be embedded through action planning (Collin and Smith, 2021). Behaviour change models such as COM-B (Capability, Opportunity, Motivation, and Behaviour) propose that capability, opportunity, and motivation are central to behaviour change (Michie *et al.,* 2011). Because each of these components interact, interventions must target one or more of these to achieve effective behaviour change. For instance, opportunity for behaviour change can be provided by ringfencing time to deliver a programme, while motivation can be enhanced by providing opportunities for professional reflection and seeing the impact.

The EEF guidance document 'Improving Mathematics in the Early Years and Key Stage 1: Guidance Report' highlights the need for early years settings and schools to invest in developing practitioners' own understanding of maths, their understanding of how children typically learn, and how this relates to effective pedagogy (Clark *et al.*, 2021). In practice, this means that settings should embed practices that support relationships with children and extend their learning, including staff knowing children well, creating a culture supporting children's curiosity, and ensuring that children engage in age-appropriate but cognitively challenging activities. Adults need to be aware of the typical development of mathematical skills and concepts to enable them to understand what may be appropriate for teaching children. Mathematical development is complex. For instance, children might be engaging successfully in certain mathematical activities despite not having a full grasp of underlying concepts. Moreover, children's development is not linear, and skills may develop in parallel (Clements and Sarama, 2018; Hodgen *et al.*, 2018).

Studies show the significant positive impact of using teaching assistants (TAs) to provide one-to-one or small group intensive support using structured interventions (Sharples *et al.*, 2021; Higgins *et al.,* 2016; Nicoletti and Rabe, 2014). Alongside working on structured interventions, the EEF guidance document 'Making Best Use of Teaching Assistants: Guidance Report' highlights the importance of TAs receiving high-quality ongoing support and training to effectively deliver the interventions (Sharples *et al.*, 2021).

There is a small but growing body of evidence to support the use of picture books and storybooks to teach maths, most of it involving the specific use of talk. The Education Endowment Foundation (EEF) evidence review of early years and Key Stage 1 maths teaching identified this as a promising approach, with six studies providing a large positive aggregate effect. Four of these studies conducted in early years settings involved the use of specially chosen or designed picture books and storybooks and involved guidance to support mathematical talk and educator questioning (Casey *et al.,* 2008; Hassinger-Das *et al.,* 2015; Purpura *et al.,* 2017; van den Heuvel-Panhuizen *et al.,* 2016). Importantly, how practitioners use picture books are key to maximising the opportunities the books provide (Clements and Sarama, 2018). The ShREC (Share attention, Respond, Expand, Conversation) approach (see 'Intervention' section below) outlines four key evidence-informed strategies that can support practitioners to have responsive and stimulating conversations with children.

Maths Through Picture Books (MTPB) is a ten-week targeted intervention that uses specific picture books to stimulate multi-turn conversations, in order to help children who are identified as being at risk of not developing a secure understanding of number before they move to Key Stage 1. Sessions are delivered to small groups of five children, thus incorporating research evidence highlighting the benefits of small group learning for children in early years settings (Wasik, 2008; Milburn *et al.,* 2014). The aim of the MTPB programme is to equip reception teachers and TAs with:

- improved knowledge of how children learn number and operations; and
- the knowledge and skills to engage a small group of targeted children in sustained multi-turn conversations about number and operations using picture books and the 'ShREC approach'.

For children, the programme aims to:

- improve enjoyment and motivation towards learning mathematics among children; and
- have children talk about number and operations during the intervention.

The MTPB programme sits within a wider focus among the previous and the current UK government on enhancing maths education. For instance, Maths Hubs were launched in 2014 to improve maths education across schools, with the previous government committing to expanding the reach of Maths Hubs to 75% of primary schools and 65% of secondary schools by 2025 (Lewis and Maisuria, 2023). The approach to maths across Maths Hubs is 'teaching for mastery'. In a typical lesson, the teacher should lead back and forth interaction, including questioning, short tasks, explanation, demonstration, and discussion, enabling pupils to think, reason, and apply their knowledge to solve problems (NCETM, 2022). MTPB's focus on interactive reading, where children engage in multi-turn conversations about numbers and operations, reflect these principles.

Additionally, the revised Early Years Foundation Stage (EYFS) framework, published in December 2023, emphasises developing positive attitudes towards maths, foundational skills in number recognition and recall, subitising, and understanding number bonds (DfE, 2023b). MTPB aligns with this framework by aiming to improve motivation towards learning maths and providing training for teachers and TAs that is focused on foundational mathematical skills (Clark *et al.*, 2021).

The EEF supported a formative evaluation of the programme's early-stage development work in 2022. The unpublished evaluation suggested that the programme is practical to implement and acceptable among participants, although the findings were not collected by an independent evaluator. Unlike the MTPB pilot evaluated in this report, it took place over seven weeks and only involved ten small schools in a rural part of North Yorkshire. Teachers and TAs enjoyed delivering the programme and liked the Continuing Professional Development (CPD) element of the programme. The programme was a positive experience for children who were reported to have built confidence and enjoyment in maths, with data suggesting that most of the children made considerable progress in a short amount of time. The evaluation concluded that it was implemented with a high degree of fidelity; however, TAs made non-permitted alterations to the intervention, such as including manipulatives[1] in the small group reading sessions. Delivery in some schools was impacted by staff and child absence, with the timetabling of sessions into the busy weeks of the Summer Term proving a challenge. It was shown to be easier to implement in larger schools due to staff availability. The formative evaluation indicated potential to scale, calling for a pilot evaluation of the programme.

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 13).

What

Procedures

In each school, one teacher and one TA is trained by the delivery team from the East London Research School (ELRS). Training involves a full-day in-person training session at the beginning of delivery and a half-day in-person training session at the end of delivery. The full-day in-person training is focused on supporting participants to understand the importance of maths in the early years and how children learn about number and operations drawing on the first two recommendations of the EEF's 'Improving Mathematics in the Early Years and Key Stage 1: Guidance Report' (Clark *et al.*, 2021). The training session also covers the ShREC approach and the importance of interactive reading for children's communication and language development, cognitive skills, and relationships with adults. The half-day in-person session revisits the content from the full-day session and focuses on practical ways of sustaining MTPB beyond the

[1] Mathematical manipulatives are artifacts used in mathematics education, enabling children to explore or investigate mathematical concepts or processes and to perform problem-solving activities. An example is plastic blocks that click together, often used to learn addition and subtraction. Images of these are included in one of the picture books and practitioners have improvised to include real-life blocks while conversing with children.

pilot. Additionally, during the ten-week intervention period, ELRS deliver an online session for teachers on how to effectively support their TAs, and an optional online drop-in session for all teachers and TAs to support delivery. Attendance at the online session for teachers is recorded by ELRS and teachers are directed to watch the recording of the session if they are unable to attend.

Guidance recommends that TAs run 20-minute sessions doing interactive reading with nine specially chosen maths-focused picture books, with one book being repeated in weeks 1 and 10. The sessions should be delivered to a small group of five children aged 4 to 5 years. The sessions, supported by semi-structured prompts for the TA, are expected to happen at the same time every week, with conversations structured using the ShREC framework. The ShREC approach developed by ELRS, aims to provide early years practitioners with a simple and memorable set of evidence-informed strategies that can be embedded into everyday practice (James, 2022). This includes:

- ***Sh*aring attention** – being at the child's level and paying attention to what the children are focused on.
- ***R*espond** – following the child's lead and responding to their non-verbal and verbal communications.
- ***E*xpand** – repeating what the child says and building on it by adding more words to construct sentences.
- ***C*onversation** – having extended back and forth interactions and giving children time to listen, process, and reply.

The teacher and TA are required to have weekly planning meetings to ensure fidelity to the programme and consistency between the targeted sessions and wider classroom practice. At each planning meeting, the TA and teacher reflect on the previous week's sessions, recording children's attendance, any challenges, and child-specific reflections. They then prepare for the following week's sessions by looking through the new book, reading the scaffold,[2] and discussing the sessions in advance. This includes identifying the mathematical focus of the sessions and using the ShREC approach to guide potential conversations, as well as discussing what the TA should intentionally focus on for each child. Finally, they are required to think through links to the classroom and opportunities to revisit, reinforce, and rehearse the mathematical focus in wider practice. The teachers can model the sessions if required. Conversations should then be recorded on a planning template and uploaded weekly to a secure Google Drive folder.

Teachers are expected to observe the TA delivering the programme three times (in Weeks 1, 4, and 10). During the observation, they are asked to use the ShREC framework and make notes in an observation template. The teachers are then instructed to have a professional conversation with the TA on the same day, providing feedback and supporting the TA to reflect on the session using the ShREC framework. The joint reflections should then be recorded on the observation template and uploaded into the school's secure Google Drive.

Materials

Schools are provided with resources to support delivery and two sets of nine books for classroom use (one for the intervention and one for use in the classroom outside of the intervention). Resources include a programme handbook, planning template, observation template, ShREC framework, video exemplification, and the picture book scaffolds. All resources, templates, and frameworks are made available in the handbook and on Padlet.

How

The intervention is delivered face-to-face to groups of five children aged 4 to 5 years, identified by teachers as requiring extra support. To select children for the pilot, the participating teacher carries out a short activity based on the Give-N task[3] with all children in one reception class. The use of the task is an adaptation for the evaluation in order to reduce subjectivity when selecting pupils. This activity enables the teachers to identify children who have not yet developed a

[2] The scaffold is a one-page document accompanying each book and included in the programme handbook, which includes suggestions of how practitioners can share attention, respond, expand, and have sustained conversations with the children.
[3] The Give-N task is widely used in the field of developmental psychology to indicate young children's knowledge or use of the cardinality principle, wherein the last number word used in the counting process indicates the total number of items in a collection (see Baroody and Lai 2022).

secure understanding of cardinality, namely, being able to count numbers in a set. Teachers are provided with a Padlet containing guidance on how to carry out the activity and an observation grid for recording the key information (see Appendix 6). Where teachers identify more than five children who find the task challenging, they are encouraged to use their professional judgement and knowledge of each individual child to select those children they believe would benefit most from the programme. It is worth noting that the selection process is different for the pilot compared to normal practice for the intervention, where teachers would select children *after* receiving the training based on their understanding of the programme, their professional knowledge of children's mathematical abilities, and any assessment data they have collected. To replicate the sequence of activities in a possible trial, teachers were asked to select pupils *before* attending the training as part of this pilot.

Where

MTPB sessions are expected to run in a dedicated, quiet space away from the classroom, and so participating schools needed to be big enough to support this. While the Memorandum of Understanding (MoU) stipulated the importance of providing space for the sessions to be run, this was not an explicit inclusion criterion (Appendix 1). To enable the in-person training to be viable and to cover a range of different schools, pilot schools were recruited from Essex and the East London boroughs of Redbridge, Havering, and Newham.

Who provided

Fliss James and Melissa Prendergast from ELRS provided the training and support for schools. Fliss and Melissa, the developers of the intervention, are highly experienced early years teachers and leaders with significant expertise in designing and delivering high-quality evidence-informed professional development. TAs were expected to deliver the intervention sessions, supported by teachers.

Research questions

The evaluation set out to answer a total of 15 research questions mapped against the EEF's three pilot criteria (see Table 2):[4]

- 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 inception phase. The research questions about feasibility reflect the causal mechanisms and contextual assumptions set out in the theory of change and associated logs. The research questions about evidence of promise address the intended and unintended consequences stated in the theory of change. The research questions about readiness for trial consider scalability and the ability of an evaluation to test the programme's theory of change, including intended impacts.

In this report, we present evidence against 15 key success indicators to inform the decision about whether MTPB should proceed to trial. The success indicators were developed in close collaboration with the delivery team 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 below (Table 2) 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 below.

[4] The order of the research questions was changed in this report compared to the research questions outlined in the study plan. In this report, we decided to cover questions on feasibility of implementation first before moving on to the evidence of promise and readiness for trial in order to ensure a better flow and coherence of the overall report.

Table 2: Research questions

Pilot criteria	Research questions
Feasibility of implementation	1. Is the programme delivered as intended? • Is training and support for practitioners delivered as intended? • Is the programme delivered as intended in school?
	2. What contextual factors affect implementation? • Are there any particular challenges around including disadvantaged children?
	3. What adaptations are participants making to the programme and why?
	4. How appropriate are the pupil selection/eligibility criteria? • How well do the selection/eligibility criteria enable schools to include disadvantaged children? • How might the timing of selecting pupils pre-testing and pre-training influence intervention outcomes?
	5. Are programme inputs (training sessions, physical resources, additional support) accessible and useful to practitioners? • How well do the inputs support practitioners to include and support disadvantaged children?
	6. Is the programme affordable for schools?
Evidence of promise	7. Is there an improved understanding, among TAs and teachers, of how children learn about numbers and operations?
	8. Is there an increased confidence in teaching maths among TAs and teachers?
	9. Do children talk about number and demonstrate increased enjoyment of and motivation towards learning mathematics? • Are disadvantaged children[1] perceived to benefit more or less than their peers?
	10. Is teaching practice embedded and sustained after ten weeks? • Is the use of picture books to teach maths sustained? • Is the use of principles around selecting picture books to teach maths sustained?[2] • Is the use of ShREC strategies sustained?
	11. Are there any unintended consequences of participation in this programme? • Are ShREC strategies used outside of the intervention? • Is the intervention perceived to have benefits for children's socio-emotional learning and/or language and communication? • Are there benefits for child–adult relationships? • Does the intervention take practitioners away from other activities and is this trade-off perceived to be worth it? • Does the intervention take children away from other activities and is this trade-off perceived to be worth it? • Do children selected for the intervention feel stigmatised? • Do children not selected feel they are missing out? • Do any unintended consequences affect disadvantaged pupils more or less than their peers?
	12. How different is the programme to business as usual?
Readiness for trial	13. Can the programme feasibly be delivered at a larger scale? • Is there a clear and realistic plan for what scaling to deliver in more schools would look like and require? • What modifications would be required to scale delivery? • How might the timing of selecting pupils pre-testing and pre-training influence intervention outcomes?
	14. How well do we understand the causal chain and mechanisms which underpin the intended effects of the programme?
	15. What potential measures could be used in a subsequent efficacy trial?

[1] The EEF focuses on socio-economic disadvantage, using free school meals (FSM) eligibility as a marker of disadvantage. However, when asking practitioners about disadvantaged pupils, we did not offer a definition when asking about disadvantage.

[2] The training did not include guidance on selecting appropriate picture books for teaching mathematics, and the programme's picture books were pre-selected and provided by the delivery team. Therefore, this sub-question was not answered in our evaluation.

Table 3: Success indicators

Pilot criteria	Success indicators (research question number)	Sources of evidence
Feasibility of implementation	Programme is delivered with fidelity (RQ 1)	• Training attendance data • Other monitoring information from Padlet • Summer Term survey (quantitative teacher/TA-reported fidelity) • Qualitative interviews with teachers/TAs (qualitative self-reported fidelity and adaptation—the how and why) • Observations (qualitative and quantitative observations of agreed core components)
	Pupil selection/eligibility criteria are appropriate and feasible to implement (RQ 4)	• Qualitative interviews with teachers/TAs (qualitative perceptions of accessibility and usefulness) • Summer Term survey (quantitative teacher/TA perceptions)
	Programme inputs (training, physical resources, additional support) are accessible and useful (RQ5)	• Qualitative interviews with teachers/TAs (qualitative perceptions of appropriateness and feasibility) • Summer Term and Autumn Term surveys (quantitative teacher/TA perceptions)
	Programme is affordable to schools (RQ6)	• Qualitative interviews with a member of the SLT (qualitative perceptions of affordability) • Qualitative interviews with teachers/TAs (qualitative perceptions of affordability of demands on time) • Summer Term survey (quantitative teacher/TA-reported)
Evidence of promise	Improved understanding among TAs and teachers of how children learn about numbers and operations (RQ7)	• Summer Term (post-delivery) and Autumn Term (longitudinal follow-up) surveys (quantitative teacher/TA-reported improvements) • Qualitative interviews with teachers/TAs (qualitative self-reported improvements)
	Increased confidence in teaching maths among TAs and teachers (RQ8)	• Summer Term and Autumn Term surveys (quantitative teacher/TA-reported increases) • Qualitative interviews with teachers/TAs (qualitative self-reported increases)
	Children talk about number and operations during the intervention (RQ9)	• Summer Term and Autumn Term surveys (quantitative TA reports on proportion/regularity of children talking about number/operations) • Qualitative interviews (qualitative TA reports on the how and why) • Observations (qualitative and quantitative observations)
	Improved enjoyment of and motivation towards learning maths among children (RQ9)	• Summer Term and Autumn Term surveys (quantitative teacher/TA reports on improvements) • Qualitative interviews (qualitative teacher/TA reports on improvements) • Child discussion groups (qualitative self-reported improvements)
	Teaching practice embedded and sustained after ten weeks (RQ10)	• Autumn Term survey (quantitative teacher/TA-reported practices)

	No indication that the programme likely widens the gap between disadvantaged pupils and their peers (RQ11)	• Qualitative interviews with a member of the SLT (qualitative perceptions of promise) • Qualitative interviews with teachers/TAs (qualitative perceptions of promise)
Readiness for trial	Core programme elements identified (RQ13)	• Intervention Delivery and Evaluation Analysis (IDEA) workshop • Desk-based review • Learning workshop
	Clear and realistic plan for what scaling to deliver in more schools would look like and require (RQ13)	• Interviews with delivery team • Learning workshop
	Plan for scaling would entail little or no modification to in-school delivery (RQ13)	• Interviews with delivery team • Learning workshop
	Causal chain and underpinning mechanisms articulated (RQ14)	• IDEA workshop • Desk-based review • Learning workshop
	Outcome measure(s) identified that is/are: • aligned to intended outcomes • able to measure changes that we expect to see, and • acceptable to schools (RQ15)	• Desk-based review • Piloting assessments (primary measure for a trial) • Child discussion groups (secondary measure for a trial)

Ethical review

The project was reviewed and approved by the Oxford MeasurEd ethics board in August 2023. 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 not involved in delivery of the project. The chair then convened a meeting with the reviewers and the project team where they discussed issues arising and agreed actions to be taken forward. The project director approved these actions and the board chair signed off the approval.

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

School agreement was obtained through an MoU (Appendix 1). The MoU was signed by the headteacher as well as the participating teacher and TA 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 participants' contact details.

Data protection

Data collected for this evaluation was used only 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 including preventative measures and processes for reporting, reviewing, and responding to breaches. All data transferred to Oxford MeasurEd for the purposes of the evaluation occurred 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. Names and associated codes were stored separately from the pseudonymised data in a data file only accessible to members of the research team. We did not report schools' or individuals' names in evaluation outputs, and considered the disclosure risks of including any quotes, but some participants may be recognisable due to the small size of the pilot evaluation—particularly the two members of the delivery team. We have communicated these limits to confidentiality and anonymity to participants before and after each data collection activity (see Appendix 7 for an example of a topic guide outlining these limitations).

For this evaluation, Oxford MeasurEd was the data controller who also processes data. This means that we were responsible for deciding the purpose and legal basis for processing data. Our legal basis for data processing was 'legitimate interests'. For any special categories of personal data (ethnic identity of participating school staff and children), we used the research exception. Data subjects were the delivery team, TAs, teachers, children, and SLT members. We outlined data protection procedures and safeguards, our legal bases for processing data during recruitment, and published a privacy notice online, which we circulated to all concerned parties.

After the evaluation ends, pupil data from the piloting of outcome measures will be matched with information from the Department for Education's (DfE's) National Pupil Database (NPD) and stored in the EEF archive, which is hosted in the Office for National Statistics (ONS) Secure Research Service (SRS). At this point, the EEF will become the data controller, and Fischer Family Trust (FFT) Education (archive manager) the data processor, of the archived data. Anonymised data will be accessible to the wider research community for secondary analyses that provide public benefit and are in line with the missions of the EEF, DfE, and ONS. Further matching to NPD and other administrative data may take place during this subsequent research. All personal data about participants will be securely deleted from Oxford MeasurEd's systems no more than six months after submission of data to the EEF data archive and the EEF final edited report.

Project team

The project was delivered by a team from ELRS, based at Sheringham Nursery School and detailed in Table 4:

Table 4: Delivery team

Name	Role	Institutional affiliation
Fliss James	Project delivery lead	ELRS
Melissa Prendergast	Project delivery lead	ELRS

The evaluation was delivered by a team from Oxford MeasurEd and detailed in Table 5.

Table 5: Evaluation team

Name	Role	Institutional affiliation (role at institution)
Dr Lydia Marshall	Principal investigator and project director	Oxford MeasurEd (research director)
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 (consultant)
Paulina Valenzuela	Junior evaluator and assessment expert	Oxford MeasurEd (consultant)
Claire McAlpine	Evaluator	Oxford MeasurEd (senior consultant)
Ross Duncan	Evaluator	Oxford MeasurEd (associate)

Methods

Recruitment

The delivery team recruited 20 schools for the pilot evaluation across four areas: Essex; Newham; Havering; and Redbridge. The delivery team made a conscious decision to focus on three London boroughs (Havering; Newham; and Redbridge) and Essex to enable them to have a range of schools but did not consider Ofsted (Office for Standards in Education, Children's Services and Skills) gradings or specific demographics. They also selected these areas as they were in close proximity to the location of ELRS, therefore, making intervention delivery such as training attendance more practical. Recruitment took place between October 2023 and December 2023. The team recruited via the EEF website, the ELRS website, targeted 'newsflashes' for schools in the four areas, social media posts, and support from colleagues in local Teaching School Hubs and Maths Hubs via newsletters and the social media. Take-up was quick and smooth, so the delivery team did not offer any other specific recruitment activities such as in-person events.[5] The chosen sample size considered both ELRS's capacity constraints, which allowed them to deliver training and support to no more than 25 schools, and evaluation requirements, which required a minimum of 16 schools to enable eight to ten evaluation interviews to be conducted without both teacher and TA at any one school being interviewed. Oxford MeasurEd, ELRS, and the EEF decided on 20 schools as the sample size to deliver best value for money.

The delivery team and the evaluation team monitored school recruitment during the recruitment phase to ensure that:

- participating schools were not participating in other maths/numeracy focused interventions in reception funded by the EEF, to avoid contamination; and
- the pilot included a diverse sample from different geographical areas, including a mix of rural, coastal, and urban schools as well as a mix of school sizes and types, including infant schools.

As part of the signing up process, participating schools were required to select a TA and teacher from the same reception class to take part in the intervention. There were no eligibility criteria for these practitioners, other than the school authorising their time to be spent attending training and delivering the intervention. The teacher, TA, and headteacher then signed an MoU agreeing to the requirements of the programme and the evaluation. The MoUs were shared by ELRS with the evaluation team via a secure platform (Google Drive) on a regular basis during the recruitment period.

To select five children for the pilot, the participating teacher at each school was instructed to carry out a short activity based on the Give-N task with all children in the same reception class (see 'Intervention' section above). The activity took place between 20 November 2023 and 15 December 2023. Once schools selected the five children for the intervention, they were requested to share a parent/carer information sheet with the parents/carers of the selected children, giving them the opportunity to withdraw their child from the evaluation (see Appendix 4).

The evaluation team prepared all recruitment materials for evaluation activities (see Appendices 1 to 4). 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 followed the EEF implementation and process evaluation (IPE) guidance (EEF, 2022) and pilot guidance (EEF, 2023) and included desk-based research, primary research with the delivery team, teachers, TAs, school SLT members, and children, and observations of MTPB delivery.

Figure 1 sets out how the design addressed the three pilot criteria and the 15 research questions.

[5] As this was a pilot evaluation, ELRS recruited exactly 20 schools and did not have a waitlist.

Figure 1: Evaluation matrix

Evaluation domain	Research question	Desk-based research			Research with DT			Research with school staff					Observations		Research with pupils	
								Interviews			Surveys					
		Programme documentation	Programmatic data	Outcome measure review	IDEA workshop	Interviews with delivery staff	Learning workshop	TAs	Teachers	SLT	Teachers	TAs	Training	Classroom	Outcome measure piloting	Pupil focus groups
Feasibility	1. Is the programme delivered as intended?															
	2. What contextual factors affect implementation?															
	3. What adaptations are participants making to the programme and why?															
	4. How appropriate are the pupil selection/eligibility criteria?															
	5. Are programme inputs (training sessions, physical resources, additional support) acceptable and appropriate?															
	6. Is the programme affordable for schools?															
Evidence of promise	7. Is there an improved understanding, among TAs and Teachers, of how children learn about numbers and operations?															
	8. Is there an increased confidence in teaching maths among TAs and Teachers?															
	9. Do pupils talk about number and demonstrate increased enjoyment of and motivation towards learning mathematics?															
	10. Is teaching practice embedded and sustained after 10 weeks?															
	11. Are there any unintended consequences of participation in this programme?															
	12. How different is the programme to business as usual?															
Readiness for trial	13. Can the programme feasibly be delivered at a larger scale?															
	14. How well do we understand the causal chain and mechanisms which underpin the intended effects of the programme?															
	15. What potential measures could be used in a subsequent efficacy trial?															

DT=delivery team.

Table 6 provides an overview of the different participant groups, the activity, the 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. For instance, while we nearly achieved our aim of carrying out ten interviews with TAs, only ten out of 19 TAs responded to the post-intervention survey. Due to the large ask of schools for this pilot evaluation (including baseline and endline pupil assessments), schools received a financial incentive of £500 for completing the IPE activities and pupil assessments.

Table 6: Overview of pilot participants and activities

Participant group	Activity	Mode and duration	N (target)
Maths Hubs representatives	Business as usual interview	Online, 30–45 mins	Two interviews (no target)
Delivery team	IDEA workshop	In-person, 120 mins	One workshop
	Paired interview	Online, 90 mins	One interview
	Learning workshop	In-person, 120 mins	One workshop
TAs	Qualitative interviews	In-person or online, 35–55 mins	Eight interviews (ten targets)
	Post-intervention survey	Online, c. 15 mins	Ten responses (19 targets)
	Longitudinal follow-up survey	Online, c. 5 mins	Five responses (19 targets)
Teachers	Qualitative interviews	In-person or online, 30–60 mins	Eight interviews (ten targets)
	Post-intervention survey	Online, c. 15 mins	16 responses (19 targets)
	Longitudinal follow-up survey	Online, c. 5 mins	Nine responses (19 targets)
	Outcome measure interviews	In-person, c. 10 mins	Four interviews (four targets)
SLT	Qualitative interviews	In-person or online, 15–40 mins	Eight interviews (eight targets)
Observations	Training	In-person (c. half-day–one day) and online (c. 30–90 mins) as per training mode	Five observations (five targets)
	In-school delivery: Intervention sessions	In-person, c. 20 mins	Ten observations (ten targets)
	In-school delivery: Planning sessions	In-person, c. 10 mins	Three observations (no target)
	In-school delivery: Reflection sessions	In-person, c. 10 mins	Three observations (no target)
Children	Outcome measure piloting	In-person, c. 10 mins (baseline) to 20 mins (endline)	Baseline: 97 children successfully tested at 20 schools Endline: 89 children successfully tested at 19 schools
	Child discussion groups	In-person, c. 25 mins	Three discussion groups (five targets)
Assessors	Outcome measure interviews	In-person or online, c. 10 mins	Four interviews (four targets)

Desk-based research

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

- understand the theory of change, including causal mechanisms and contextual assumptions;
- map research evidence collected against the causal mechanisms and contextual assumptions so far; and
- identify and decide on programmatic data to be used for the evaluation.

We also carried out a desk-based **review of potential outcome measures** for a future trial. We agreed the methodology for this review with the EEF in April 2023. Our review included considering, for each potential outcome measure:

- the alignment between what they measure and the intended outcomes of MTPB;
- their statistical properties;
- logistical factors such as mode of data collection, length of assessment, and any training/briefing required to administer the assessment; and
- their suitability for use with reception children.

We analysed the pros and cons of each identified measure and outlined these for discussion with the EEF and the delivery team. Following the review, we identified the Wechsler Individual Achievement Test-Third Edition UK (WIAT-III UK) as an appropriate measure for testing children participating in the MTPB programme. A note outlining the methodology and findings of the review can be found in Appendix 5.

We also completed a light touch review of **other early years maths provision in the pilot areas** to understand business as usual (BAU) and the context in which pilot schools are implementing the intervention, primarily by looking at school and local authority websites and extracting relevant information. We complemented the desk-based research with 45-minute conversations with key informants from Maths Hubs in two pilot areas.

Finally, we analysed programmatic data to understand attendance rates at training and the extent to which practitioners completed and uploaded their logs. We were unable to analyse how TAs and teachers used Padlet, as it was not possible for ELRS to monitor usage (e.g. number of downloads and views of documents) on Padlet.

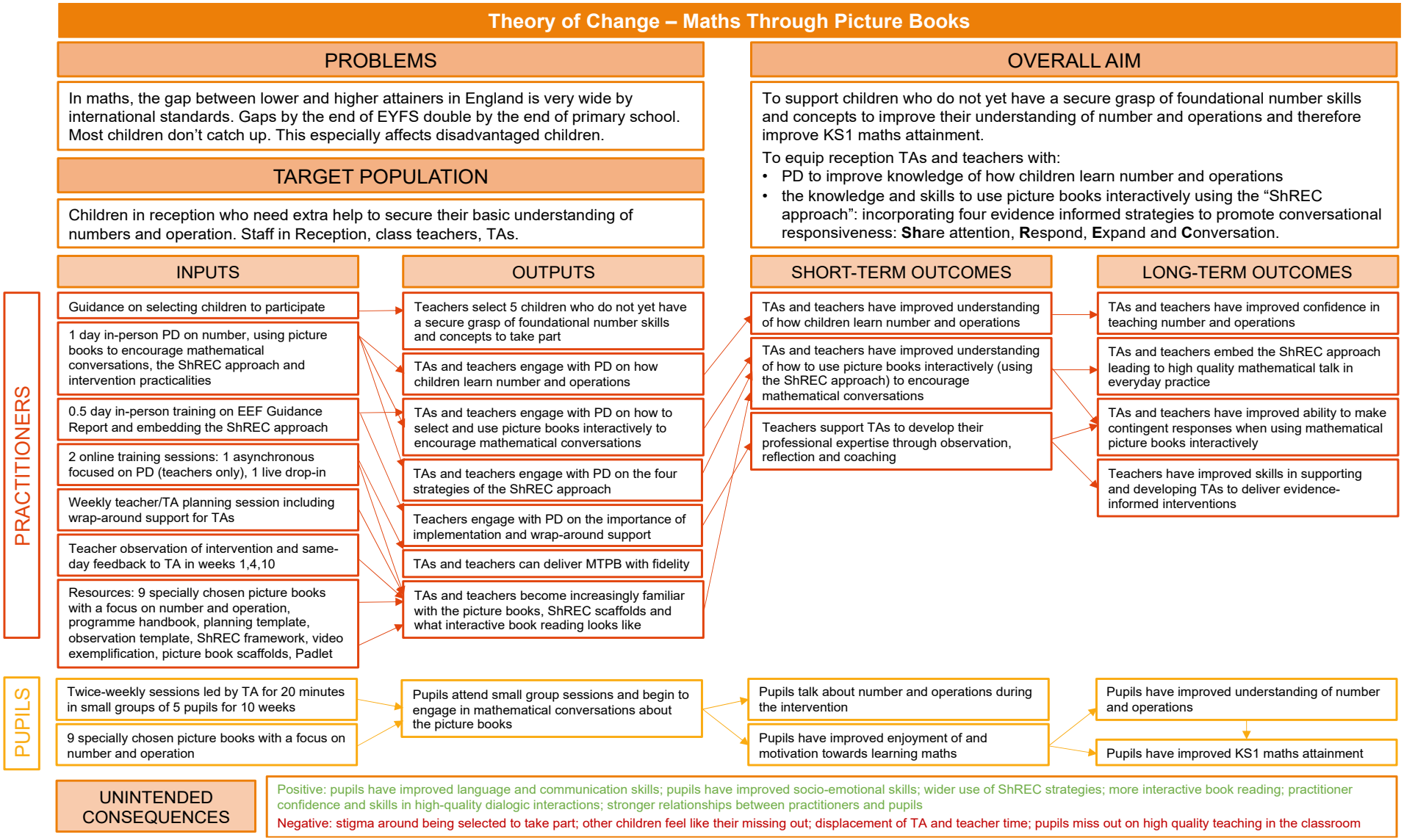
Research with delivery team

We facilitated an in-person **IDEA workshop** with the delivery team in February 2023, following the EEF guidance. In this workshop we discussed the theory of change, assumed causal mechanisms and contextual assumptions, defined compliance and fidelity, and identified priorities for the pilot evaluation to explore in terms of fidelity and adaptation. Following the workshop, we finalised the theory of change (see Figure 2) in close consultation with the delivery team and the EEF.

We carried out a **90-minute paired qualitative online interview** with both members of the delivery team in June 2024. The interview gathered their views on the feasibility of the training, the differing training needs of teachers versus TAs, evidence of promise in terms of TA and teacher understanding of the MTPB methods, and reflections on scalability.

We also facilitated a two-hour in-person post-delivery learning workshop with the delivery team in July 2024 to explore lessons learned by the team, discuss the delivery team's plans for scaling delivery, and seek participant validation of our interim findings.

Figure 2: Theory of change (developed by the delivery team with facilitation from the EEF and the evaluation team)



KS1=Key Stage 1; PD=professional development.

Research with school staff

Qualitative interviews with school staff

We carried out qualitative interviews with school staff (TAs, teachers, and SLT members such as heads of school, deputy heads, or Foundation Stage leaders) between March 2024 and June 2024 to explore questions related to evidence of promise, feasibility, and readiness for trial (see topic guide example in Appendix 7). The overall aim was to carry out a total of 28–30 interviews across TA, teachers, and SLT members in order to capture the full range of experiences and views across the sample and to account for the range of characteristics within our sample. Overall, we completed 24 interviews with 25 individuals[6] from pilot schools, including nine teachers, eight TAs, and eight SLT members. Of these, 16 interviews were online, with the remaining eight interviews taking place in-person. Before each interview, we underlined that only the research team had access to the data and provided reassurances around anonymity. At the end of each interview, we checked in with participants to ensure they were happy for everything that they discussed to be included in the report. The interviews lasted on average around 45 minutes, although there was some variation in interview length, for example, due to participants' other commitments. To ensure consistency in data collection and minimise bias a senior member of staff quality assured the first interview each interviewer had with a member of school staff.

Quantitative surveys of school staff

We invited all teachers and TAs to complete a short online post-delivery survey shortly after the end of delivery in June 2024 to quantify their experiences and views on the programme's evidence of promise and feasibility, and perceptions of benefits and unintended consequences. The post-delivery survey also asked about intentions to continue with the MTPB intervention, using the ShREC strategies, and using picture books to teach children about number, and about other early years maths interventions that the school had engaged with that year. Overall, 16 teachers and ten TAs successfully completed the post-delivery survey.

We also carried out a second **online longitudinal follow-up survey** of teachers and TAs in October 2024 to capture self-reported sustained practice. Overall, only nine teachers and five TAs successfully completed the longitudinal follow-up survey, despite several targeted reminders sent to encourage their participation. Due to the small sample size, we only draw on the longitudinal follow-up survey findings when reporting on sustained practice.

To minimise bias and reassure participants, we reminded participants that only members of the research team would have access to the survey responses and that there were no 'right' or 'wrong' answers.

Observations of training and in-school delivery

We also carried out observations of training and in-school delivery between March 2024 and June 2024. The primary aim of the observations was to provide insights to tailor the prompts and probes used in interviews with the delivery team, teachers, and TAs. We gathered observational data on teachers and TAs' responsiveness to training, children's responsiveness to MTPB sessions, any challenges to successful delivery and strategies to overcome these, and any adaptations to delivery.

We observed the full-day in-person training session in March 2024 and half-day in-person training session in June 2024. We also observed an online training session for teachers as well as an optional live drop-in session for teachers and TAs, both in March 2024. We aimed to observe ten intervention sessions and successfully carried out these observations at ten schools between March 2024 and May 2024. At three of these schools, we also observed a planning session between teachers and the TA. At another three of these schools, we observed a reflection session between the teacher and TA following an observation of the session by the teacher.

[6] In one SLT interview, a deputy head and an early years maths lead were interviewed together.

Research with children

Qualitative research with children

We carried out three child discussion groups in May 2024, each of which took place after the observation of the intervention session. They took place in-person and lasted around 25 minutes. The group discussions were attended by two members of the evaluation team to ensure quality assurance and consistency in data collection and to better manage group dynamics. The purpose of the discussions was to explore how children felt about being selected for the intervention, their views on the session content, and the extent to which they enjoyed and were motivated towards learning maths.

Our approach drew on good practice guidance (e.g. see Brady and Graham, 2019), for instance by keeping activities interactive and as brief as possible. We also consulted and asked ELRS to review the topic guide as experts in both the intervention and in working directly with this age group.

The original aim was to carry out a total of five child discussion groups. However, because of challenges in getting meaningful data, we decided with ELRS and the EEF not to continue with this activity after completing the third discussion group. While children talked about enjoying the stories and pictures, we did not elicit more detailed findings. One reason for this was the short period of time we spent at the school, making it difficult for children to establish rapport with the researcher. For this reason, our reporting does not draw on any findings from the child discussion groups to make an assessment on the degree to which the criteria have been met. Appendix 8 includes a more detailed discussion of our approaches and reflection on some key lessons.

Piloting of outcome measure collection

We collected baseline *and* endline outcome data for 89 children participating in the pilot using the WIAT-III UK, the assessment we identified in our desk-based review of outcome measures as the best option for a future trial of MTPB. We collected outcome data using the WIAT-III UK Maths Problem Solving subtest at baseline in January 2024 and February 2024 and outcome data from the WIAT-III UK Maths Problem Solving and Numeracy subtests at endline in June 2024.

The assessment was administered digitally using tablets by a team of freelance teachers managed by Oxford MeasurEd. The assessment was administered one-on-one with the child and each subtest took approximately ten minutes to complete.

A member of the Oxford MeasurEd evaluation team familiar with the assessments carried out spot checks to quality assure the data collection. This involved visiting two schools at baseline and another two schools at endline to ensure the assessments were carried out as intended and to a high standard. A different assessor was observed at each spot check.

As part of the spot checks, we carried out short ten-minute in-person or online interviews with teachers and the assessors we observed. The aim was to understand the acceptability of the process and how it might be improved, to identify any key risks and challenges, and to explore how schools might best be incentivised to use this measure.

Data analysis

We recorded workshops and interviews with participants' permission, using the 'Framework' approach to manage 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, meaning that the analysis was in part deductive, and in part inductive. This enables 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 (StataCorp LLC, College Station, Texas, USA). We agreed a detailed analysis plan with the EEF.[7] The evaluation director 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.

We analysed quantitative data from the outcome measure piloting to provide key statistical information to inform the design of a future trial (e.g. pre- and post-correlation, intraclass correlation coefficients [ICCs], and data missingness). The statistical information was analysed in Stata 18 and included the following:

- Assessment of the pre- and post-outcome distribution to assess for floor and ceiling effects through histograms.
- Reporting the mean, standard deviation (SD), and median of the measure at baseline and endline.
- Estimation of the Pearson correlation between pre- and post-outcomes.
- An estimation of how much variance in the post-outcome is explained (R^2) by the pre-test using multi-level linear regression model. The dependent variable will be the post-outcome, regressed against the pre-outcome, with a random intercept for providers.
- The ICC was estimated from the multi-level linear regression model with the post-estimation command *estat icc*.
- Attrition was assessed by the proportion of pupils who completed a pre-test who did not take part in the post-outcome testing.
- Item missingness was assessed through tabulations.

We did not analyse the data to measure changes in pupil outcomes over the course of the pilot since: i) we had a small sample size (maximum n=100); and ii) we did not have a counterfactual estimate of the progress children would make in the absence of the intervention. Our impact evaluation expert quality assured the outcome measure analysis syntax and output.

Our analysis was ongoing throughout the pilot. 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 our evaluation director. 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 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 recruitment period, data collection, and delivery schedule.

[7] The study plan, including the analysis plan, can be found **here**.

Table 7 Timeline of activities

Date	Activity
February 2023	IDEA workshop with delivery partners
February 2023 – March 2023	Set-up meetings
February 2023 – April 2023	Desk-based review of programme documentation
February 2023 – May 2023	Desk-based review of outcome measures
March 2023 – July 2023	Study plan
April 2023 – July 2023	Recruitment materials
June 2023 – August 2023	Ethical approval
July 2023 – December 2023	Instrument development
September 2023 – October 2023	Desk-based review of business as usual
October 2032 – December 2023	School recruitment and teacher and TA enumeration
November 2023 – December 2023	Pupil selection
January 2024	Pupil enumeration
January 2024	Pupil baseline assessments
March 2024 – June 2024	Training
March 2024 – June 2024	Training observations
March 2024 – June 2024	Collection of programmatic data
March 2024 – June 2024	In-school delivery
March 2024 – June 2024	Delivery team staff interviews In-school observations School staff interviews Child discussion groups
June 2024 – July 2024	Post-intervention surveys of TAs and teachers
June 2024 – July 2024	Pupil endline assessment
July 2024	Learning workshop with delivery team
August 2024 – October 2024	Final analysis of main evaluation data
September 2024	Interim findings presentation for the EEF grant committee decision-making about proceeding to trial
October 2024 – November 2024	Longitudinal follow-up survey of TAs and teachers
November 2024	Longitudinal follow-up survey analysis
December 2024	Final draft report

Findings

Participants

A total of 20 state schools from four areas in the Greater London and East London area participated in the pilot. This included eight schools from Essex, six schools from Newham, and six schools from Redbridge and Havering. Three of the participating schools were infant schools, and the remaining 17 were primary schools.[8] One school withdrew after baseline testing due to staff changes, with the remaining 19 schools remaining in the evaluation throughout the pilot.

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

Area overview

Although there are areas of high deprivation in **Essex**, residents report higher earnings than the national average, as individuals commute to better-paid jobs (Department of Education, 2017). Essex contains a larger proportion of people describing themselves as 'White British' (85%) than the national average (74%) (Essex County Council, 2021).

Havering is characterised mainly by suburban development, with relatively low levels of deprivation in comparison with other London boroughs (London Borough of Havering, 2023). Havering is also one of the most ethnically homogeneous boroughs in London: the ethnic make-up is predominantly White British (75%), with an increasing population of Asian, Asian British, and Asian Welsh residents (ONS, 2021a).

Newham is characterised by densely populated residential areas (ONS, 2023) and is the third most deprived borough in London (Newham Council, 2022). Newham is the most ethnically diverse local authority in England and Wales, with 69% of residents identifying as non-White (ONS, 2022). Residents are more than four times as likely to identify with 'Asian' or 'Black' ethnic groups than the England average.

Redbridge is a borough of contrasts, with 11 neighbourhoods that rank among the 20% most deprived in England, and another 11 among the 20% least deprived (London borough of Redbridge, 2024). Redbridge is the third most diverse local authority in England and Wales, with a White population of 35%. Around 47% of residents identify as 'Asian', 'Asian British', or 'Asian Welsh' (ONS, 2021b).

School and pupil profile

Participating schools did not have a 'typical' profile but displayed variation regarding the type of school, their Ofsted rating, the number of reception classes, class size, and the overall profile of their reception cohort.

Nine out of the 19 pilot schools (47%) were local authority-maintained schools (also known as community schools), while seven pilot schools (37%) were academies (academy converters or academy sponsor-led). Three schools (16%) were either voluntary controlled or voluntary aided schools.[9] This broadly reflects the current national picture of school types among primary settings.[10]

Thirteen schools (68%) have been rated 'good' by Ofsted, with six schools (32%) having been rated 'outstanding'. Thus, a higher proportion of pilot schools have been judged 'outstanding' compared to the national picture, where 12% of primary schools have received this grade (EPI, 2024).

The number of reception classes in schools ranged from one to four. There was less variation in terms of class sizes, with most participants reporting classes to include around 30 children. The lowest number of children reported among teachers, TAs, and SLT members who commented on their school's profile was 22 children. No participants reported class sizes higher than 30 children.

[8] An infant school is an educational institution that caters to young children, usually between reception and Year 2 (ages 4 to 7). They are relatively uncommon nationally, with the majority of 4- to 7-year-old pupils attending primary school for children aged 4 to 11.
[9] Voluntary aided and voluntary controlled schools are faith schools. However, unlike voluntary aided schools, voluntary controlled schools are solely funded by the local authority.
[10] For instance, 42.7% of primary schools in England are currently academies or free schools (DfE, 2024a).

The make-up of the reception cohort tended to mirror the socio-economic and ethnic make-up of the wider area. In some areas, participants highlighted an increase in newly arrived families and pupils from Ukraine, Afghanistan, and China, as well as more established immigrant populations from the Americas, the Philippines, West Africa, Eastern Europe, and South Asia. This meant that schools, particularly in more urban and densely populated areas, had a sizeable number of children with English as an Additional Language (EAL) in their reception cohort.

Levels of disadvantage among the reception cohorts were similarly varied. The sample included schools with universal or near-universal FSM eligibility to schools where only one in 20 pupils qualified as socio-economically 'disadvantaged' based on being eligible for FSM.

A recurring theme among participants was an increase in the proportion of children with Special Educational Needs (SEN), with participants citing figures as high as one in three pupils in some reception classes having SEN status and having an Education, Health and Care Plan (EHCP). There was a view that this was partly driven by the Covid-19 pandemic, with pupils in reception classes displaying less advanced social skills than pre-pandemic cohorts.

School staff participant profile

Table 8 provides a breakdown of all programme participants based on a short enumeration form completed by teachers and TAs in December 2023. One teacher and three TAs did not complete the enumeration form. Their background characteristics are recorded as 'unknown'.

Table 8: School staff participant profile

	Teachers		TAs	
	No.	%	No.	%
Area:				
London boroughs	11	57.9	11	57.9
Essex	8	42.1	8	42.1
Experience:				
Four years or less	6	31.6	11	57.9
More than four years	11	57.9	5	26.3
Unknown	2	10.5	3	15.8
Education:				
Below NVQ Level 5	0	0	7	36.8
NVQ Level 5 and above	18	94.7	9	47.4
Unknown	1	5.3	3	15.8
Age:				
40 or less	12	63.2	7	36.8
Over 40	6	31.6	9	47.4
Unknown	1	5.3	3	15.8
Ethnicity:				
White	15	78.9	9	47.4
Non-White[1]	3	15.8	7	36.8
Unknown	1	5.3	3	15.8
Total	**19**	**100**	**19**	**100**

[1] *This includes Asian or Asian British, Black, Black British, Caribbean or African, Arabs, and mixed ethnic groups.*

Among programme participants, there were notable differences regarding their professional experience between teachers and TAs. While 58% of all teachers had more than four years' experience, most of the TAs (58%) had four years or less of experience in their job. All teachers were university educated, with a small proportion also having a master's degree. In contrast, only 47% of TAs had a qualification above National Vocational Qualification (NVQ) Level 5 and a small proportion did not have any higher education qualifications. All age groups were represented among teachers and TAs. A larger proportion of the teachers were aged 40 or younger (63%). In contrast, a larger proportion of TAs (47%) were over 40 years old. The participants identified overwhelmingly as female.[11] A higher proportion of TAs (37%) than teachers (16%) identified as non-White.

Achieved sample

Research with school staff

Qualitative interviews with school staff

Teachers

The primary criterion for selecting teachers was geographical area. We set a quota for each region based on the percentage of schools from each region in the overall population to capture variations in experiences, perspectives, and contextual factors across different locations. In addition to the primary criterion of area, we considered four secondary criteria to capture the diversity of experiences of the teachers facilitating the programme: years of experience; age; gender; and ethnicity of the participants. The aim was to ensure that the initial sample broadly reflected the overall profile of the teacher population. While we achieved this for nearly all characteristics, this was not the case for ethnicity. Almost all teachers we interviewed identified as White, even though 16% of participating teachers did not identify as White.

TAs

Our aim was to include either the TA or teacher at each of the 20 participating schools. Since we had background information available for more teachers, we selected teachers to be interviewed first and then selected TAs for interviews at those schools where we did not plan on interviewing teachers, monitoring for diversity throughout. Our interviewed sample of TAs broadly reflected the overall profile of TAs participating in the programme.

Quantitative surveys of school staff

Post-delivery survey

Teachers

The response rate for teachers was 84% (16 out of 19), and we achieved a good level of diversity in terms of the teachers' experiences including their years of experience, age, gender, and ethnicity, with no patterns in non-response.

TAs

The response rate for TAs was lower at 53% (ten out of 19), despite targeted reminders sent to encourage their participation. Although we achieved diversity in the range of experiences among TAs, non-White TAs from London boroughs were less likely to respond to the survey.

Longitudinal follow-up survey

Teachers

The response rate for teachers was lower than for the post-delivery survey at 47% (nine out of 19). Younger teachers with fewer years of experience were less likely to respond to the longitudinal follow-up survey. Some non-responses may be due to the possibility that teachers had left the school by the time the longitudinal follow-up survey was

[11] Exact figures have been suppressed to reduce the risk of disclosure.

conducted; however, we did not collect data on whether teachers were still employed by the school at the time of the longitudinal follow-up survey.

TAs

The response rate for TAs was very low, at 26% (five out of 19). As with teachers, younger, less experienced TAs were less likely to respond to the longitudinal follow-up survey. As with teachers, non-responses from TAs may be because they left the school by the time of the follow-up; however, we did not collect data about TAs' employment status at the school when we administered the longitudinal follow-up survey.

Observations of training and in-school delivery

The primary criterion for selecting schools for intervention session observations was the geographical area. We set a quota for each region based on the percentage of schools from each region in the overall population to capture variations in experiences, perspectives, and contextual factors across different locations. In addition to the primary criterion of area, we considered two secondary criteria (where available) to capture the diversity of experiences of TAs delivering the programme and how differences in professional experience and educational background might shape delivery and programme outcomes: **years of experience** and **highest level of qualifications**. Half of the original sample included TAs who had not provided this information; it was important to include them in the sample, as their non-response could have indicated lower levels of motivation towards the programme. Table 9 includes a breakdown of schools by area where we carried out observations of intervention sessions.

Table 9: Achieved sample by area for intervention session observations

Area	N	%
Newham	3	30
Essex	4	40
Redbridge	2	20
Havering	1	10
Total	**10**	**100**

Research with children

Piloting of outcome measure

At baseline, 97 out of 100 children were tested across a two-week testing period. Assessors were expected to carry out a minimum of two visits to ensure as many children were tested as possible. Those children that were not tested were continually absent. Children who were not tested at baseline were not eligible for endline testing. Following baseline testing, one school withdrew from the programme and the evaluation due to staff changes, which left 19 schools in the pilot for endline testing. Across these schools, 89 out of the remaining 92 children were tested at endline. Children who were not tested at endline were either continually absent or had moved school.

Feasibility

The following section explores questions around the feasibility of implementation, covering research questions 1 to 6.[12]

Research question 1. Is the programme delivered as intended?

Key findings

Overall, training and support for practitioners was delivered as intended, with training attendance high except for drop-in sessions. In-school delivery was mainly as intended. Planning sessions tended to take place weekly and observations sessions at three specific timepoints, although there was variation in the quality of feedback offered by teachers. Intervention sessions took place twice a week in nearly all schools, with most sessions lasting 20 minutes.

The following section discusses fidelity in relation to training and ongoing support as well as in-school delivery.

Overall, training and support for practitioners[13] was delivered as intended. Training attendance was high except for optional drop-in sessions.

Training attendance at the in-person sessions was high. All 19 teachers and 17 out of 19 TAs attended the first full-day in-person training. Attendance rates for the half-day in-person training were slightly lower, with 17 of 19 teachers and 15 of 17 TAs attending (Appendix 12 Table 1).

The implementation support session for teachers was similarly well attended, with 17 out of 19 teachers present. However, attendance at the two optional drop-in sessions aimed at teachers and TAs was significantly lower, with three participants at each session (Appendix 12 Table 1).

Interviews highlighted that the limited attendance of the drop-in sessions related to capacity and need. For instance, the delivery team explained that the scheduling of the first drop-in session for 10:00 a.m. meant that many TAs could not attend. However, several teachers explained that they did not need any additional support and so decided not to attend.

Observations of training delivery for practitioners showed fidelity to the programme, with minor deviations in the structure of delivery. The duration of sessions at the full-day in-person training differed slightly from the initial agenda. For example, the introduction and evidence were covered more quickly than intended, leaving more time to spend on interactive reading and the ShREC approach. Similarly, during the half-day in-person training, the introduction and evidence sections were covered in more detail, but less time was spent on embedding maths activities in other areas.

More significant adaptations related to the number of drop-in sessions and the introduction of a weekly delivery log are discussed under research question 9 below.

The programme was broadly delivered as intended in schools. All but one schools remained in the pilot throughout the duration of the programme, meaning that 95% of pilot schools completed the programme. Planning sessions tended to take place weekly and observations sessions at three specific timepoints, although there was variation in the quality of feedback offered by teachers. Intervention sessions took place twice a week in nearly all schools, with the majority of survey respondents suggesting the sessions lasted 20 minutes.

Planning sessions generally took place on a weekly basis as intended, with 22 out of 26 respondents from the post-delivery survey reporting that the planning sessions 'took place weekly' (Appendix 12 Table 2).

Observations of planning sessions and interviews indicated that teachers and TAs followed all elements when reviewing the previous session but did not cover all elements consistently when 'planning ahead'. When reviewing the previous session, teachers and TAs reviewed attendance and discussed what went well and less well. In one school, this involved the teacher and TA starting off by watching the recording of the previous session, which was an adaptation to delivery

[12] We have decided to present findings on feasibility first before covering questions related to evidence of promise (as outlined in the study plan), as we find this to be a more logical sequence when presenting findings. We amended the research question numbers in this report to reflect this.
[13] Where we use the term practitioners in this report we refer to teachers and TAs.

(see research question 3 below). Teachers and TAs also debated specific challenges the TA faced and how to address these. This included child-specific challenges, such as proposed measures to ensure certain children were less dominant during sessions.

There was variation across the observed planning sessions regarding the detail and content covered when looking ahead at the following session. Not all practitioners read and engaged with the scaffold provided for each book, although they all identified the maths focus. Similarly, there was inconsistent application of the ShREC framework to guide potential conversations. Links to the wider classroom and curriculum were made sporadically. Interviews and observations did not provide any evidence of teachers modelling, although this was an optional element of the sessions. However, interviews and observations demonstrated that teachers and TAs consistently discussed what the TA would be focusing on for each child. For instance, a TA described how they and their teacher discussed, which children would be able to achieve the learning goals of the book and how they would help them reach these goals.

> *We are able to get to the crux of the matter [the learning goals] very quickly…We know the people, so we know who is going to get to the goal quickest or what way they are going to get there.* (TA)

Schools did not consistently upload the planning templates after each session as required. On average, schools submitted the planning templates 77% of the time based on programmatic data collected by the delivery team. The delivery team also explained during the learning workshop that the level of detail provided was inconsistent across schools. To ensure regular submissions and sufficient detail, they introduced weekly logs shortly after delivery began, which consisted of a list of mainly yes/no questions akin to a 'checklist' to ensure that schools completed all the required activities. However, the delivery team acknowledged that the weekly logs were not submitted or used consistently by pilot schools.

A total of 18 out of 26 respondents completing the post-delivery survey indicated that feedback and coaching sessions (referred to as **'reflection sessions'** hereafter) 'always' took place at the three timepoints (Weeks 1, 4, and 10), with the remaining respondents saying they 'sometimes' took place or did not take place at all (eight out of 26) (Appendix 12 Table 3).

A total of 15 out of 24 respondents reported that the reflection session always took place on the same day as the teachers' observations. Others noted this was the case only sometimes or never (nine out of 24). While we only observed three reflection sessions, these all happened on the same day as the teacher's observation.

As with planning sessions, observations of the reflection sessions showed variation across schools regarding, which prescribed elements were covered. For instance, teachers did not consistently cover all sections in their discussions with the TA. For instance, during our observations, guiding questions including: 'Was the book used as planned?' and 'Were adaptations made for children's developmental levels?' were not covered when providing feedback. The feedback often lacked detail, as it was not specific. (see research question 5 below).

Observations highlighted variation in teachers supporting TAs to reflect on their sessions using the ShREC framework. For instance, during an observation the teacher asked the TA to share their reflections without referencing ShREC strategies; this resulted in the TA's reflections being focused more on maths and group dynamics and less on the ShREC strategies.

Submission rates were slightly higher than with planning templates; the delivery team's programmatic data shows that on average schools submitted the reflection templates 86% of the time after the reflection sessions.

Turning to **intervention sessions** for children, post-delivery survey respondents were asked how often two MTPB intervention sessions took place each week throughout the ten-week pilot, with response options including 'every week', 'at least every second week', and 'fewer than every second week'. They were also asked whether these sessions consistently lasted 20 minutes, with options of 'no, they were less than 20 minutes', 'no, they varied', and 'yes, they were always 20 minutes long'. A total of 15 out of 26 respondents reported that MTPB sessions consistently occurred every week and always lasted 20 minutes. Among the remaining 11 respondents, while some noted inconsistencies in session

frequency, most indicated that sessions generally took place weekly but varied in duration, sometimes lasting longer or shorter than 20 minutes[14] (Appendix 12 Table 4).

Observations of intervention sessions as well as interviews with teachers and TAs broadly echo the survey findings. Sessions tended to be 20 minutes long, although in some schools they were shorter or longer. Where TAs delivered shorter sessions, this was either due to children getting increasingly distracted or because of misunderstandings; for instance, in a session, the TA incorrectly assumed that they had to read half the book on one day and the other half on the second day of the week.

Interviews with teachers, TAs, and the delivery team as well as intervention session observations highlighted that intervention sessions were nearly always delivered by TAs. Where teachers delivered the sessions rather than TAs, this was either due to TAs being unavailable or teachers modelling certain strategies to support their TA.

MTPB sessions tended to take place in a dedicated quiet space away from the classroom, with 22 out of 26 survey respondents noting that this was 'always' the case. The remaining four reported that sessions only 'sometimes' or 'never' took place in a dedicated quiet space away from the classroom (Appendix 12 Table 4). Illustrative of schools that were unable to provide a dedicated quiet space was a school we visited during an observation; here, a session took place within an open space between two classrooms in which people walked across the corridor, which occasionally distracted the children.

Where dedicated quiet spaces were used, this included different types of rooms such as libraries, designated intervention rooms, or other enclosed spaces outside classrooms. A group of schools used the same location each week, while others had to compromise and used different spaces. Other schools purposefully switched between different locations to give children some variety.

TAs generally used ShREC strategies when delivering MTPB sessions, although there was variation in the quality of delivery (see research question 5 below). Overall, 22 out of 26 survey respondents reported that they 'always' delivered the sessions using the ShREC approach. The remaining four respondents reported they used ShREC 'sometimes' or 'never' (Appendix 12 Table 5).

Participants that did not consistently follow the ShREC approach found it challenging to avoid asking many questions (see research question 5 below). For instance, a TA reflected how they used loaded questions to help children unpack concepts:

> *You're not supposed to ask questions, but if I can see somebody's mind unpacking an idea or unpacking a concept a certain way…I have to ask a loaded question or two to get them to continue to unpack it.* (TA)

Others established explicit rules around reading, contrary to programme guidance. For instance, a TA explained how they had told children at the beginning of the session to take turns, as strategies such as moving the book around had resulted in some children not getting a chance to speak during previous sessions.

Across all observations TAs used the right mathematical picture book with a focus on number and operations. A TA explained that they displayed the relevant book for the week in the classroom, as they wanted children to become familiar with it and children not selected to take part to be exposed to the books too.

Interviews and observations of intervention sessions highlighted that TAs generally did not use any other resources in the intervention other than the book. In the few instances where TAs used resources, these included props (e.g. introducing crabs for a crab book) or using painted numbers on the classroom wall to point out numbers to children.

[14] Exact figures have been suppressed to reduce the risk of disclosure.

Research question 2. What contextual factors affect implementation?

Key findings

SLT members and teachers were key in facilitating delivery, with teachers playing an important role in supporting TAs. A perceived lack of time and a lack of quiet dedicated space were key factors, which complicated delivery for some schools. Moreover, there was a view that it could be challenging to deliver the programme to children with mixed abilities as it required the TA to tailor the session to a range of different needs.

The following section discusses contextual factors affecting implementation with reference to the implementation environment and/or implementer factors. Broadly speaking, the implementation environment refers to contextual characteristics in schools where MTPB took place, while implementer factors pertain to the profile of professional characteristics and attitudes among implementers (Humphrey e*t al.,* 2016). This is discussed in relation to: i) training and ongoing support, including from the delivery team to participants and from teachers for TAs; and ii) intervention sessions.

Factors affecting the implementation of training and support

Implementation environment

Participants discussed several interrelated factors affecting training and/or support. These included SLT support, staffing, and time.

Participants cited SLT support as particularly beneficial for attending training and carrying out planning and particularly reflection sessions. Seven out of 25 survey respondents agreed that SLT support made it easier for them to access the training and support provided by the delivery team (Appendix 12 Table 8), while five out of 24 survey respondents noted that support from SLT members made it easier to carry out reflection sessions (Appendix 12 Table 12). There was no evidence from the post-delivery survey and the interviews that a lack of SLT support was a barrier to implementation.

SLT support revolved primarily around ensuring that cover was available for participants engaging in MTPB-related activities. Participants explained during interviews that they found it reassuring that their SLT members had organised cover when they attended training or took part in planning and reflection sessions. They also highlighted how SLT members protected participants' time in advance, so that their time would not be taken up by other commitments. For instance, a teacher described how all the programme activities—including the training and ongoing support—were incorporated into a timetable so that they could be delivered as intended:

> *We do kind of run a timetable. It's not just all free flow, so I think that's probably why it fitted in quite easily to what we're doing*. (Teacher)

Staffing challenges mainly related to schools struggling to find cover to enable practitioners to take part in MTPB activities. Some interviewed teachers and TAs recognised that it could be challenging for both to attend training, especially where schools needed to source external cover (see research question 6 below). For instance, a teacher explained that they could not attend the second online drop-in session as no cover had been available on the day of the session.

A perceived lack of time constituted a key barrier to accessing and/or engaging with training and ongoing support. A total of 11 out of 26 survey respondents believed time made it difficult to access the training and support provided by the delivery team (Appendix 12 Table 7). Moreover, 17 out of 26 survey respondents believed a lack of time made it more difficult to carry out the planning (Appendix 12 Table 9) and 15 out of 24 survey respondents reported the same for reflection sessions (Appendix 12 Table 11).

The limited time available was a key reason for the limited attendance of drop-in sessions, as teachers and TAs described during interviews how other job-related commitments meant they were unable to attend. It also translated into participants spending less time engaging with the resources available and fully consolidating the learning they gained from the training sessions. Furthermore, there were instances where other time commitments meant teachers and TAs could not plan together; an SLT member therefore, questioned whether their teachers had time to facilitate the programme by supporting their TA through planning and reflection sessions:

> *It's just the time element. The teachers, they did a fantastic job, but it's just: Have they got the time to facilitate it?* (SLT member)

Implementer factors

Support from teachers played a significant role in facilitating TAs' participation in planning sessions. Interviews with TAs underlined that teachers were available for TAs, especially when TAs were unclear about something. This collaboration suggests that TAs valued the activity being carried out together.

> *We go through the book and have a chat, but then also I do see [teacher] after school and after the session [planning session], if something's happened…I can talk to [teacher] at any sort of time.* (TA)

Some TAs highlighted that they had known the teacher for years and had strong professional relationships. They believed this made it easier for them to approach teachers and ask for support.

> *I know her so well, so I think if there's been any issues…I can message her.* (TA)

The delivery team acknowledged that TAs with less experience of CPD were likely to have found the programme more challenging, as they would not be used to analysing and reflecting on their behaviour the way it was required by the programme. Similarly, the delivery team explained that TAs with limited knowledge and experience of early years pedagogy were likely to have experienced more challenges with implementation, because the programme was targeting skills that those TAs were unlikely to possess, including actively developing relationships with children and conversing in a less structured way:

> *Relationships with children and to have conversations is not something that they would normally be asked to do. Teaching assistants who are not used to early years practice are going to be used to taking children out for interventions that are very formal.* (Delivery team interview)

Factors affecting the implementation of intervention sessions

Implementation environment

Teachers played a crucial role in helping TAs deliver sessions. This support was reflected in teachers actively encouraging TAs to deliver their sessions and protecting their time.

> *[Teacher] knows it's every Monday afternoon and every Thursday afternoon that I need to step out of the classroom and come into the library with the five children. And as soon as it's time, it's just 'take the children away. Don't worry about whatever you are doing'.* (TA)

SLT members tended to play a more hands-off role with overseeing intervention sessions. In interviews, participants explained that SLT members being invested in the programme and showing 'respect' meant that it was easy for them to book spaces for the intervention sessions.

Teacher or TA absences in some schools had implications for the delivery of intervention sessions and required flexibility around timings or sessions needing to be rescheduled. For instance, a school reported being unable to release the TA for a session when the teacher was unwell, as this would have affected the staff–child ratio.

Nine out of 26 survey respondents highlighted a lack of time as a barrier to delivery (Appendix 12 Table 13), suggesting that while this was a barrier for some, a lack of time was not a barrier for the majority of survey respondents. In qualitative interviews with teachers and TAs, there was a general awareness of teachers and TAs having limited time available, but this was not reported as a barrier to delivery. Instead, they emphasised how SLT and teacher support (e.g. by protecting time) ensured that a lack of time did not become a barrier to delivery.

During interviews with TAs, participants highlighted the benefits of having a quiet, dedicated space to carry out the MTPB sessions. They believed a quiet and dedicated space made children more focused as well as making the session less formal, for instance because it made it easier for TAs to sit on the floor. Schools that struggled to find a quiet space either had building works at the school or limited availability of quiet spaces overall. Delivering the sessions in a space that was not quiet meant children got easily distracted and impacted the quality of the session.

Participants reported that the composition of the group of children impacted delivery. Qualitative interviews revealed that some participants reported challenges when working with children that had different abilities. There was a view that it could be challenging to deliver the programme to children with mixed abilities as it required the TA to tailor the session to a range of different needs, which could be overwhelming.

> *Sometimes it's overwhelming and they've all got different abilities and [different] things that I'm focusing on, I can't even think straight.* (TA)

Other participants found it challenging to deliver the session in groups where there was a mix of dominant and quiet children (see research question 4 below). They used strategies such as increased eye contact to engage quieter children, although this was not always effective.

The delivery team explained that schools' business as usual approach to play-based learning likely shaped participants' experience of implementation and the quality of delivery. They highlighted that certain schools experienced a contrast between the programme's approach to interactive reading via the ShREC approach and the more didactic and structured approach evident in some schools.

> *Some…classes are so formal that there is very little room for play and very little room for…responsive conversations.* (Delivery team interview)

The delivery team's reflection was mirrored in interviews with teachers, TAs, and SLT members. TAs unfamiliar with the ShREC approach or those previously encouraged by their school to ask many questions struggled with ShREC strategies.

> *We're so used to asking them questions all the time.* (TA)

> *She's found it quite difficult not to ask the questions as such and to let the children feed her. But as it's gone on, it's got easier, and the children have got more of the idea of how it works.* (Teacher)

There is no indication from the quantitative and qualitative findings that there were particular challenges to including disadvantaged children. Challenges around including certain pupil groups (who may or may not be disadvantaged) are discussed in research question 4 below.

Research question 3. What adaptations are participants making to the programme and why?

Key findings

Adaptations to training and support for practitioners were minor and included setting up a WhatsApp group to facilitate communication among practitioners, extending the number of drop-in sessions from one to two, and introducing a delivery log to encourage settings to submit their planning and reflection templates. Adaptations to in-school delivery included some TAs videoing themselves delivering the session or schools using the books outside the intervention session.

Adaptations to training and support for practitioners

The delivery team introduced three minor adaptations once delivery had started. These included the number of drop-in sessions, the use of WhatsApp, and the weekly delivery log.

The delivery team had initially planned on having one online drop-in session to reduce the burden on schools. They decided to increase the number of drop-in sessions to two shortly after the first session had been completed. This was due to several reasons: the scheduling of the first session for 10:00 a.m. in the morning had made it difficult for many participants to attend; there was limited interaction and engagement on WhatsApp (see paragraph below), specifically among TAs; and the planning and reflection logs included limited detail. The second drop-in session took place after school hours between 4:00 p.m. and 4:30 p.m., although this did not improve attendance and the same number of participants (three) attended.

The delivery team decided to set-up a WhatsApp group shortly after the programme began in order to facilitate communication between participants and for participants to ask each other and the delivery team questions.

The delivery team introduced a weekly delivery log a few weeks into delivery in response to several schools either not submitting the planning or reflections templates on time and/or the templates lacking detail. The purpose of the log was to function as a 'checklist' to ensure schools submitted all the relevant information and to make schools more accountable to complete the templates. Questions included: whether the planning/reflection session took place; whether the planning log was submitted; whether the observation was completed (if applicable) and the reflection log submitted; what book was used; and whether the session took place in a quiet space. The delivery team explained that they had used the weekly logs during the pre-pilot stage but then decided to initially scrap it before the pilot to reduce the burden on schools.

Adaptations to in-school delivery

Adaptations to in-school delivery were made to: i) planning and reflection sessions; and ii) intervention sessions.

Planning and reflection sessions

Observations of planning and reflection sessions and qualitative interviews pointed to the use of video recordings as the only (permitted) adaptation made to planning and reflection sessions. Where TAs video recorded their intervention sessions, it was unclear if they had intended to do this from the outset or not, as video recordings were not part of the intervention as such. A TA explained how they recorded themselves delivering every intervention session, which they used when meeting the teacher for planning and reflection sessions. They considered recordings helpful to remind them how the session went and believed it made the process of reflecting on the session easier. Our observations of planning and reflection sessions and interviews with TAs, teachers, and SLT members did not provide any evidence of non-permitted adaptations.

Intervention sessions

Key adaptations observed during intervention sessions and discussed during qualitative interviews related to pupil selection as well as the use of books and other equipment.

Some schools with more than one reception class selected children from two or three reception classes rather than one, as advised by the delivery team. A teacher explained that this allowed selection to be driven by need, although it required more teachers to be involved in selection and more logistics, for instance because every child completed the Give-N task (see research question 4 below).

A group of teachers and TAs explained that they used the book outside the intervention session, which was a permitted adaptation highlighted in the training slides. Examples of what this looked like in practice included using the book with the whole class and recapping the book with any children who were absent during the weekly sessions.

Another adaptation that related to materials included the use of whiteboards to facilitate delivery. For instance, a TA explained how they used the whiteboard to remind them what to focus on during the sessions but made sure it was not visible to pupils.

Research question 4. How appropriate are the pupil selection/eligibility criteria?

Key findings

All teachers who completed the post-delivery survey confirmed that they followed the guidance from the delivery team on how to use the Give-N task. Overall, they regarded the task as useful and relevant, but expressed more mixed views about whether they selected the right children for the task. The timing of selecting pupils before testing and training led to some schools selecting pupils who they might have not selected with a better understanding of the programme. Moreover, there was a view that the guidance could have been more explicit about the suitability of the programme for children with SEN and EAL.

This section outlines the process of selection within schools, including who carried out the task and the different criteria involved in decision-making. It then explores participants' views on the usefulness of the task before and whether the right pupils were selected.

Selection process

All teachers who completed the post-delivery survey confirmed that they followed the guidance from the delivery team on how to use the Give-N task (Appendix 12 Table 15). The guidance stipulated that teachers needed to carry out the activity with *all* children in their reception class and, where more than five children could be selected to take part in the intervention, they should use their professional judgement and knowledge to make a final selection. Qualitative interviews with participants showed variation in TAs' involvement in decision-making. In several schools, teachers carried out the task themselves without involving TAs. In the cases where schools selected children from several reception classes, selection involved multiple teachers. In other schools, teachers sought their TAs' views once they had selected the five children.

Usefulness and relevance of task and guidance

Participants' views on the usefulness and/or relevance of the guidance and task were largely positive. Seven out of the 16 teachers responding to the post-delivery survey agreed that the guidance for child selection was 'very useful', with the remaining nine teachers reporting that it was 'useful' or 'somewhat useful' (Appendix 12 Table 16). During qualitative interviews, participants described the process of administering the task following the guidance as straightforward and comparable to the Reception Baseline Assessment (RBA). They found it beneficial that it provided them with an additional objective assessment of pupils' mathematical abilities that they could use beyond the intervention and for the whole class. They also liked that it was relevant to the intervention, as the task assessed what was taught through the mathematical picture books.

> *You know what you're getting…is what they're sort of teaching through the books and what the aims of the lessons are.* (Teacher)

However, schools that assessed all children in the cohort rather than assessing children from one reception class described the task as more 'arduous' than expected. A less pronounced view was that the task did not offer any additional insights and teachers could have selected the right children without carrying out the test.

> *I would have been able to tell you who could have done [the programme].* (Teacher)

Interviews with teachers and SLT members revealed two similar but different approaches to selecting pupils using the task. The first approach consisted of selecting those children with the lowest scores on the test without using additional criteria to inform selection. A second group of teachers followed the guidance by using their knowledge of children to select the children. Some schools used inclusion/exclusion criteria to guide decision-making, including:

- **Fit between pupil and intervention.** An example included a teacher, in consultation with the TA, selecting those children from the longlist who would be most content with having their routine disrupted by being taken out of the classroom for the MTPB intervention sessions.
- **Confidence with maths.** An example included a school basing their selection partly on how confident they thought children felt about maths.
- **Potential benefits.** This involved selecting those children who were most likely to benefit and/or excluding those who would not. Specific examples included:
    - excluding pupils with SEN, for instance because they were already receiving targeted support through an EHCP or because teachers deemed their verbal and social skills insufficient to benefit;
    - excluding high achieving pupils, even where they did not do as well as expected in the task;
    - including 'late' joiners, so that the programme could help them 'catch up'; and
    - including disadvantaged children, for instance targeting children from a specific low-income neighbourhood to take part in the programme.
- **Group dynamics.** An example included a teacher considering a range of pupil characteristics to ensure a diverse group of pupils, leading them to select a child with EAL, a quiet child, and a more dominant child.

Suitability of pupils

Even though participants generally found the task useful, they expressed mixed views on whether they selected the right children for the programme. Ten out of the 16 teachers who responded to the post-delivery survey agreed that '*all*' the children selected were right, while the remaining six teachers reported that 'some' of the selected children were right (Appendix 12 Table 17). Almost no survey respondents cited poor pupil selection as a barrier to delivery.[15] Teachers, TAs, and SLT members who believed they chose the right pupils cited a range of reasons for this during interviews when reflecting on pupils' participation in the programme. One view was that the selected children were suitable because they made the desired and expected progress in maths since taking part in the programme. Others noted that they had selected the right children because the group included those who needed and had benefited from additional support in language and communication development alongside maths, both of which they believe the programme targeted:

> *Not only were they the ones who struggle with some concepts of maths...a couple of them are those that really want that adult attention and time to talk and time to listen.* (SLT member)

Other participants cited the group composition as evidence that they had selected the right mix of children, as their group included the right mix of talkative and quieter children.

Participants who believed they did not select the right pupils noted several reasons. There was a view that some groups could have included those children who had done well on the test but would have benefited from more support. In such cases, pupils' test performances were not seen as indicative of their actual maths ability. Other groups included children who participants thought were too advanced in maths but were included because they had not performed well on the test.

> *There is one who is in the programme who is quite strong. I feel like maybe he shouldn't have [been in the programme, but he] didn't do as well in the assessment.* (SLT member)

Other participants believed they could have considered more criteria when selecting pupils. For instance, a teacher suggested that maths anxiety and conversational ability would have been useful criteria to draw on when selecting pupils, as this could have better predicted their enjoyment and attitudes towards maths.

> *That way we might be able to pick not just based on their knowledge, but also how they perceive maths and how they get on with and how they enjoy maths.* (Teacher)

Certain children being disruptive during sessions was another reason participants gave when reflecting on why certain pupils should not have been selected. There was a view that they could have paid more attention to underlying needs, especially where children did not have a diagnosed SEN.

> *So, for two of the children that we've chosen to do it…we shouldn't have really chosen them because I think they didn't get as much out of it that other children would, but also their learning behaviour is such that they disrupted…some of the sessions for the other children…If we were to do it again, we would pick the cohort of children to go through a little bit better.* (SLT member)

Schools generally included disadvantaged[16] children, although there was a view that the guidance could have been more explicit about the suitability of the programme for children with SEN and EAL (who may or may not be disadvantaged).

Qualitative interviews with teachers, TAs, and SLT members highlighted that schools generally included disadvantaged children in the group but interpreted disadvantage in different ways. One approach was to include geographical disadvantage as a selection criterion by targeting children for selection from a specific geographical area of disadvantage. More commonly, however, schools did not include disadvantage as a specific criterion but ended up including children that mirrored and/or had slightly higher levels of disadvantage among the reception or school cohort.

[15] Exact percentage has been suppressed to reduce the risk of disclosure.
[16] As previously mentioned, the EEF focuses on socio-economic disadvantage and views FSM eligibility as a proxy of disadvantage. However, we did not specify in our interviews and surveys what we meant by disadvantage.

For example, a SLT member explained that one in three children at their school were eligible for Pupil Premium while three out of five children in the MTPB class were classed as disadvantaged. Another SLT member from a school in a more affluent area noted that there were fewer than one in ten children at the school who were disadvantaged, which was reflected in the make-up of selected children:

> *That group was representative of our kind of our demographics. So not, yeah, we don't really have a great deal of disadvantaged children.* (SLT member)

A group of participants highlighted the inclusion of children with SEN and EAL within their group. Examples included all or nearly all selected children having EAL as well as children that were in the process of receiving an EHCP. However, some teachers communicated during interviews that they could have included more children with EAL or SEN but were either unclear how much they would benefit from the programme or had concerns that other learning barriers would have made programme participation difficult.

The delivery team accepted that the guidance could have been more explicit about certain pupil groups (see research question 13 below). However, they explained that some schools likely had a 'deficit' view of children with EAL and SEN and deemed these pupil groups unsuitable for the intervention:

> *There were questions from schools about including children with SEN. They felt that it wasn't appropriate because their expectations of what their children can achieve was very low…Some schools [felt] that perhaps [children with EAL] wouldn't get very much out of it because they've got to concentrate on learning English...So, I think that there is a bit of a deficit idea about SEN children or children who are learning English.* (Delivery team interview)

The timing of selecting pupils before testing and training led to some schools selecting pupils who they might have not selected with a better understanding of the programme. It is possible that children who were not selected might have benefited more from this programme than other children who were selected pre-training.

Some participants wanted pupil selection to take place after training. This was because better knowledge of the training would have resulted in them selecting slightly different children and—by implication—ensured better intervention outcomes.

> *You kind of get much more of an insight into [MTPB], when we did the selection, it was kind of before anything and you didn't really kind of understand how the project was going to work.* (Teacher)

A teacher mentioned the example of maths anxiety as something they did not consider at selection but would have considered as a selection criterion after having attended the training.

> *Since they [the delivery team] were talking about, you know, having maths anxiety…me and my [TA] did say that possibly it might have been better to have the training and then do the assessments so that we know more in depth what the training and what the sessions look like.* (Teacher)

The delivery team acknowledge a risk that without suitable guidance and information about the programme, maths subject knowledge, and group dynamics, the selection of pupils before training could lead to schools selecting the 'wrong' pupils. They acknowledged that this made it more important for the delivery team to carefully consider how much and what information to share with participants before training.

Research question 5. Are programme inputs (training sessions, physical resources, additional support) acceptable and appropriate?

Key findings

Participants were highly positive about the training and support from the delivery team. Those participants unfamiliar with the ShREC approach before the intervention found it useful learning about it during the training. Teachers liked being able to attend the in-person training with their TAs as it ensured a shared understanding. However, participants had reservations about the timings of the implementation support session and online drop-in sessions. Participants' views about programme resources provided by the delivery team were broadly positive. TAs highlighted the support received from teachers as useful, while reflection sessions with teachers provided an opportunity to learn from

teachers about areas of improvement. During sessions, TAs largely used language that was pitched at the right level for reception-aged children, and children appeared to enjoy the sessions. We did not find any evidence of disadvantaged children being excluded from programme activities.

This section outlines participants' views and our observations on the acceptability and appropriateness of programme inputs, focusing on training and ongoing support from the delivery team, and support from teachers and intervention sessions.

Training and ongoing support from the delivery team

Participants were highly positive about the training and support from the delivery team. All 26 survey respondents who completed the post-delivery survey found it 'very useful' or 'useful'. Similarly, 17 out of 26 survey respondents thought that the delivery team made it easy for them to access the training and support (Appendix 12 Table 6). Qualitative interviews with teachers, TAs, and SLT members provided insights into what usefulness looked like in practice during the training and throughout delivery.

Quality of training and support

Participants praised the quality of the training, and highlighted the clarity and detail provided. For instance, a teacher explained that everything was explained clearly from start to finish and described the training as *'some of the best training that I've had'* (Teacher). Others highlighted that the training included a lot of practical examples and trainers made enough time to elaborate on and explain everything covered, including during lunch breaks. This had the effect of participants feeling reassured before delivery.

> *I didn't feel nervous starting something completely new because they really spoke through it thoroughly.* (TA)

Participants highlighted the quality of support during delivery and found the delivery team accessible and proactive. By regularly checking in on participants, the delivery team made them feel they were always there to help, even if no help was needed.

> *There's always someone there ready to help.* (TA)

The quality of training was mirrored in how participants felt about planning and delivering the sessions after attending the training. Respondents were asked how confident they felt about planning the intervention sessions and how much they understood about how to deliver the intervention sessions after the initial training and support received from the delivery team, with options including 'not at all', 'somewhat', 'quite a bit', and 'a lot'. All survey respondents said that they felt 'a lot' (11 out of 26) or 'quite a bit' confident (15 out of 26) about planning MTPB sessions after the training (Appendix 12 Table 18). Similarly, all survey respondents agreed that they understood how to deliver MTPB sessions 'a lot' (19 out of 26) or 'quite a bit' (seven out of 26) (Appendix 12 Table 19). There were slight but notable differences between teachers' and TAs' views about planning and delivering the sessions, however. Teachers were more likely than TAs to feel highly confident about planning sessions following the training. Likewise, teachers were more likely to believe they fully understood how to deliver the sessions compared to TAs.[17] It is possible that differences in teachers' and TAs' confidence and comprehension relate to many TAs' inexperience with CPD programmes, early years maths, and early years pedagogy, as highlighted by the delivery team during the interview.

Teachers' confidence in carrying out observations and sharing feedback with TAs was more mixed. Teachers responding to the survey were asked how confident they felt about observing the intervention sessions and giving feedback after the initial training and support received from the delivery team, with options including 'not at all', 'somewhat', 'quite a bit', and 'a lot'. Half of the teachers who completed the survey (eight out of 16 respondents) reported feeling 'somewhat' confident after their online session. The remaining half felt either 'quite a bit' or 'a lot' confident, with most leaning towards only being 'quite a bit' confident[18] (Appendix 12 Table 20). While qualitative interviews with teachers did not indicate that they lacked confidence, the delivery team highlighted that the implementation support

[17] Exact figures have been suppressed to reduce the risk of disclosure.
[18] Exact figures have been suppressed to reduce the risk of disclosure.

session lacked video exemplification and felt that this could have made teachers feel more confident about carrying observations and delivering feedback.

Mode

Teachers and TAs highlighted the value of in-person training during interviews. The opportunity to meet and network with fellow teachers and TAs was a key attraction of the in-person training.

> *What I found really beneficial was there was [sic] a lot of like-minded people kind of putting their heads together.* (TA)

Participants highlighted several other benefits of in-person training, including:

- **Increased engagement.** Enabling participants to ask questions and have conversations that cannot be replicated online.
- **Peer learning.** Enabling schools to talk to other schools about their implementation plans and how best to address shared challenges, leading to 'a sense of community'.
- **Local connections.** Allowing schools to forge connections with other local schools that could be beneficial and sustained beyond the intervention.

A group of participants enjoyed the combination of in-person with online training, as it gave schools more flexibility to release teachers. However, there was a view that online training could be less engaging, with a teacher describing the implementation support sessions as less useful because it was easier to 'switch off'. Online training also limited opportunities to practice; for instance, the delivery team acknowledged that the implementation support session did not provide enough opportunities to model reflection and feedback, which they believed would have been easier to do in-person.

Co-attendance

Teachers liked being able to attend the in-person training with their TAs because it ensured that they 'were on the same page from the beginning'. One teacher decided to bring their TA to the implementation support session, as they thought it would be beneficial for them to engage with and understand the evidence around behaviour change. The delivery team noted that having the teacher and TA present at all sessions had the advantage of teachers not being relied on to cascade the information to TAs (see research question 13 below).

Content

Participants were positive about the training content. Qualitative interviews surfaced several features that participants found particularly useful.

Videos of example sessions had the benefit of enabling practitioners, especially TAs, to understand how to carry out the sessions in practice and what to avoid.

> *Having that as an example...gave me a bit of understanding of how to carry out the session and how not to carry out the session.* (TA)

> *Often you have training, and it can be very abstract [but]…everything was very explicit about how this was put into practice.* (Teacher)

The delivery team highlighted the importance of some of the video examples being delivered by actual TAs. This made it easier for TAs to realise that delivering the session was feasible.

> *Seeing someone who has the same role and type of qualification as you is very empowering and makes it more accessible and relatable and reduces the feeling that it is a long way to go or that it is not possible.* (Delivery team interview)

Those participants who had been unfamiliar with the ShREC approach before the intervention found it useful learning about it during the training.

> *[ShREC is a] slightly different way of teaching to what I've been used to for years. So that was really interesting.* (Teacher)

The delivery team explained that this perception related strongly to practitioners wanting to teach early years maths in a less rigid and formal way (see research question 12 below). In comparison to many schools' business as usual, ShREC facilitated teaching that was developmentally appropriate, playful, and with obvious links to the EYFS.

> *A lot of reception teachers want to feel like what they're doing is developmentally appropriate…they don't feel like they're doing things that are too formal, too soon, too rigid. When it comes to maths, there's a real sense that maths is abstract, it's complex. The subject knowledge and the pedagogy can be quite formal and detached from the other aspects of early childhood education and care. It can often be taught in a way that feels uncomfortable and anxiety-provoking. It sometimes feels like it's just not aligned with the rest of the early years foundation stage.* (Delivery team interview)

During interviews, some TAs acknowledged however that learning about ShREC was a 'learning curve'; this was especially the case when they had different expectations about what the programme was about. For instance, a TA explained that they had assumed the programme would simply be about using picture books with numbers and had not appreciated that this was complemented by a specific approach to interactive reading.

As with ShREC, there was a view that the mathematical concepts learned during the training were directly relevant to practitioners' day-to-day practice. For instance, a TA found the information about mathematical developmental progressions[19] as outlined in the EEF guidance report (Clark *et al.*, 2021) informative, as it helpfully unpacked the different developmental steps children went through and not to make assumptions about what they knew.

> *It really made me open up my eyes into some of the things you assume the children would know…it dissected down maths.* (TA)

Teachers explained that they enjoyed hearing about the evidence underpinning the programme, as it provided a rationale for why it could prove promising. The delivery team noted that it was important to begin with what the evidence says is impactful to persuade participants to think about changing their behaviour.

Timings

Participants had reservations about the timings of the implementation support session and online drop-in sessions. For instance, a teacher suggested that there could have been a time gap between the initial in-person session and the implementation support session, as some of the information had been covered at the full-day training and was seen as repetitive. Moreover, the delivery team suggested that one hour was not enough for teachers to fully engage with and absorb all the content around behaviour change, especially for teachers from schools whose business as usual to teaching reception-aged children contrasted with the responsive conversations encouraged by the MTPB programme.

The delivery team indicated that the scheduling of the first drop-in session for 10:00 a.m. was one of the primary reasons for poor attendance (see research question 1 above). They suggested a time shortly after school finished would have been more suitable. This was echoed by participants who cited other commitments during the drop-in sessions as a reason for not attending the drop-in session. However, attendance at the second drop-in session remained the same (n=3) despite being held after school hours between 4:00 p.m. and 4:30 p.m.

Resources

Participants felt broadly positive about programme resources provided by the delivery team. In qualitative interviews, teachers and TAs highlighted the accessibility and appropriateness of resources. Participants appreciated being pointed to other resources during training, for instance on Padlet. This gave them the opportunity to consolidate their knowledge and read about other relevant areas without being overwhelmed with information.

[19] Mathematical development progressions describe how individual skills or concepts develop over time. While there is generally an order in which these skills may emerge, development does not take place in clearly defined linear steps. Some children may develop several skills at the same time.

> *[I]f you want to research further and it's kind of something I've gone away and had a little bit more of a look at…So that…was one pretty useful area.* (Teacher)

Even though participants did not always submit the planning and reflection templates on time or added sufficient detail, qualitative interviews indicated they found the templates straightforward and easy to follow. This was especially because they could be completed digitally.

Interviews with teachers and TAs revealed that they liked having access to a copy of a book before delivery, which allowed them to practice the sessions beforehand, for example, with their own children or grandchildren. Observations of planning sessions made it clear that having two copies of the book was useful, as teachers and TAs could flip through the pages individually. Participants also enjoyed having access to a physical copy of the programme handbook, which served as a refresher and reference point throughout delivery.

> *We are using those [the handbook] kind of a bit like our Bible to go alongside the project.* (Teacher)

Findings from the post-delivery survey painted a slightly more mixed picture about participants' views about programme resources provided by the delivery team. While 18 out of 26 survey respondents believed that the availability of resources helped with accessing training from the delivery team (Appendix 12 Table 6), only slightly more than half of survey respondents (14 out of 26) agreed that the resources helped with planning (Appendix 12 Table 10). The same proportion (13 out of 24) agreed that it helped with reflection (Appendix 12 Table 12). Around 14 out of 26 survey respondents also noted that resources facilitated delivery (Appendix 12 Table 14).

Peer support

Engagement with the WhatsApp group set-up by the delivery team for peer support was mixed. While all 19 teachers had signed up to the WhatsApp group, only 15 out of 19 TAs had signed up (Appendix 12 Table 21). During interviews with teachers and TAs, some participants regarded the WhatsApp group as less useful and highlighted limited interaction as a reason for this. However, others saw it as a useful source of information for practitioners who were struggling or who wanted to share feedback on what books worked well and less well.

Support from teachers

Planning sessions

During interviews with TAs, they highlighted the support received from teachers as particularly useful as a way of compensating for their limited experiences of planning or owning an intervention.

> *I didn't know anything about how to [plan]…because I was unaware…[H]aving that…to help...I appreciated it.* (TA)

TAs cited weekly planning meetings with teachers as a useful mechanism to mitigate against the limited experience they had delivering such a programme. However, a view expressed by some teachers during interviews was that planning sessions—in particular the process of completing the planning templates—could be repetitive.

Observations of planning sessions highlighted variation in the quality of teachers' support: some did not probe sufficiently or took on a more dominant role in the planning session at the TAs' expense.

Teachers generally guided the TAs in reflecting on the previous week's sessions. They did this through:

- open-ended questions, for example, asking the TA an open-ended question about challenges faced during the previous week; and
- probing questions, for example, asking TAs to think about the reasons why a child was not speaking.

However, teachers did not consistently ask open-ended or probing questions, such as seeking specific examples or reasons for why children improved or why some things did not go well.

TAs offered detailed reflections on previous week's sessions. For instance, a TA identified what went well for each child (such as a child counting to a big number), and what did not (struggle with one-on-one correspondence), as well as

overall for the session (e.g. using fewer questions). However, there were other instances where TAs did not reflect on the needs of all children in their group or what they themselves did well or less well.

TAs were sometimes passive during planning sessions. This was partly because teachers did not consistently encourage TAs to contribute ideas and suggestions. Although not mentioned explicitly during any interviews with participants, it is possible that TAs' more 'passive' behaviour reflected power differences between the two roles, which the delivery team identified as a barrier to participation for TAs.

Reflection sessions

TAs found reflection sessions with teachers useful, as the sessions provided an opportunity to learn from teachers about areas of improvement, such as not asking too many questions. The sessions also reminded TAs that they were already doing a lot of things 'right', which helped boost their confidence.

> *It's been really good for her to realise all the [good] things she is doing…[TA] is not the most confident in delivering the session, so there's a lot of boosting to happen there.* (Teacher)

There was a view expressed by a teacher during an interview that it was particularly important to have an observation and reflection session early on during delivery. This made it easier for teachers to support TAs during subsequent planning meetings, as these were less 'abstract' once teachers had the opportunity to observe delivery in practice.

As with planning sessions, observations of reflection sessions illustrated variation in the quality of support teachers provided to TAs. The variation in quality was particularly apparent when teachers provided feedback to TAs. Teachers provided non-specific and incomplete feedback in places, with several items not discussed that were recommended to feature as part of the professional conversation with the TA (see research question 1 above). For instance, teachers omitted certain guiding questions in the template, such as reflecting on the strategies used and the extent to which they were successful. Another example included a teacher not commenting on whether the book was used as planned or making little reference to the plans. However, there were instances of teachers discussing certain elements of ShREC; for instance, a teacher stated that the TA allowed conversations to happen and invited children through open-ended questions.

Teachers provided constructive feedback to TAs. In practice, this meant focusing on specific issues that were observed in the session as well as pointing out what went well and areas of improvement. However, teachers did not consistently encourage TAs to contribute their thoughts during the reflection session. While there were instances of teachers explicitly asking for the TAs' reflections on what they would do differently, the non-specific and incomplete feedback provided meant that TAs had limited opportunities for reflection; for example, during a session, a teacher gave limited feedback on areas of improvement, making it difficult for the TA to offer ways to improve. In another example, teachers did not probe to get more detailed reflections from TAs.

Intervention sessions

Quality

Observations of the intervention sessions revealed that TAs largely used language that was pitched at the right level for reception-aged children. This included demonstrating new words such as 'flipping' through hand gestures and unpacking more complicated mathematical content. For instance, in a session, the mathematical content of the book was advanced, including decomposing the number 11 into several smaller numbers. The TA was aware of this and tried to break it down for the children, spending time patiently and using fingers on all the children's hands to get to the final number. There were instances where *TA*s missed opportunities to explain potentially new words like 'striped', 'scales', and 'fantailed', although this did not appear to impact children's understanding and the mathematical conversations taking place.

TAs generally invited all children in the group to take part in conversation. While it was common that some children were chattier than others, TAs tended to encourage quieter children to participate. In one session, for example, a TA was conscious of a quieter child and asked them what they observed, inviting them to count objects, and holding out the book in front of them to allow time for them to observe specifically. However, there were also instances where TAs overlooked certain children. In one example, children were seated around a big table and the TA did not invite those sat

furthest from the table to come closer to the book and comment on it. It was only when children got up and approached the book of their own accord that they were able to observe closely and comment.

During intervention sessions, it was observed that a few TAs found the need to 'multitask', i.e. using ShREC, following the children's cues, having mathematical conversations and trying to keep all children engaged, challenging. They described feeling 'overwhelmed' and having 'self-doubts' when delivering the sessions, which was amplified when children were loud and overly eager. This meant that they did not always notice when children got distracted, which impacted the quality of the session.

Children's engagement

Observations of intervention sessions indicated children's enjoyment of the sessions. TAs' use of ShREC strategies as well as picture books were key factors shaping children's engagement with MTPB.

TAs used a range of strategies that children enjoyed, including:

- **Following children's lead.** Children enjoyed being able to lead and direct the conversation, for example, when sharing comments on the fish or counting the fish dots on the page.
- **Using hand gestures**. These captured children's attention, such as using hand gestures to support the text, with some children mimicking the TA's gestures of the fox moving and knocking on the floor.
- **Making personal connections.** Children liked opportunities to make personal connections with the book in response to questions such as, 'what's your favourite animal' or 'how many people are in your family'.
- **Questions inviting children's reflections.** Questions like 'I wonder how many on the next page' intrigued children and increased their focus.
- **Extending conversations.** TAs captured children's interest by using their own observations to extend conversations, including through comments and questions such as 'there is a hand on this page' or 'Did anyone notice the anchor?'.

It was evident during intervention session observations that children found the books highly engaging. They particularly enjoyed the pictures and identifying objects not visible at first sight, such as snails or spots on the ladybird. Participants echoed this during interviews.

> *I hadn't really thought about how many books there are, with numbers in, that would be so valuable.* (SLT member)

Not all books were equally appealing or engaging for children, and participants noted that some were too complex. For instance, a teacher highlighted that certain books required more explanation because they included American English words such as stroller or mailbox, which children did not know.

Our findings did not find any evidence of disadvantaged children being excluded from programme activities. There was no evidence from observations of intervention sessions or qualitative interviews that inputs prevented disadvantaged children from being included. Observations revealed that certain pupil groups participated and were included by the TAs. For instance, during a session, a child with EAL commented on the pictures in their own language; the TA then picked up on their comments and stated what they were pointing towards (the trees) in English. In another example, a child whose speech was less developed as their peers was encouraged by their TA to observe and comment.

Research question 6. Is the programme affordable for schools?

Key findings

The programme was fully resourced, and so very few participants reported incurring additional expenses such as cover costs. SLT members had mixed views about potential programme costs. While there was variation in the duration of planning and reflection sessions, most teachers and TAs reported spending an average of between 15

minutes and 30 minutes per planning and reflection session. SLT members broadly saw the time commitment required as reasonable.

This section outlines participants' views on the costs and time associated with the programme.

Costs

Schools were not expected to incur additional costs as the programme was fully resourced, and few participants reported incurring additional expenses in the post-delivery survey.[20] Among those who did, the costs included cover costs, travel to the training venue, and some additional resources for provision, although these were not required.

SLT members had mixed views about potential programme costs. The delivery team calculated the cost of delivering the programme to be £58 per pupil over three years,[21] and we put this cost to SLT members for their views. A recurring view expressed by the SLT members suggested programme costs were affordable, with SLT members describing them as 'cheap', 'reasonable', and 'good value for money'. This was partly because trained staff could train others in delivering the programme. They also thought the programme was more affordable compared to other reception-based programmes they had taken part in, although it was not clear if these were targeted or whole-class interventions.

> *[£58 per pupil over three years] is like nothing…I myself used to work on an intervention and that was like thousands.* (SLT member)

Other SLT members were more sceptical about programme costs and highlighted budgetary constraints as a reason why they needed to see clear evidence of impact before committing to any programme such as MTPB in the future.

> *It's quite a lot of money for something that's not so tangible. You'd have to look at it over a longer period of time.* (SLT member)

SLT members who had needed to bring in external staff to cover TAs and teachers attending in-person training saw the programme as less affordable because of the cover costs incurred. They thought the costs outweighed the benefits because the programme only targeted a select group of children.

> *We would have to think very carefully if we wanted to spend over £1000 on a programme that's only going to impact 6 children.* (SLT member)

Time

While there was variation in the duration of planning and reflection sessions, most teachers and TAs reported spending an average of between 15 minutes and 30 minutes per planning and reflection session (Appendix 12 Table 22 and Table 23).

Around 18 out of 26 survey respondents reported spending 15 to 30 minutes on planning sessions (Appendix 12 Table 22), while 19 out of 24 reported spending the same amount of time on reflection sessions (Appendix 12 Table 23). Others spent less than 15 minutes and some more than 30 minutes, with respondents more likely to report spending less than 15 minutes.[22]

SLT members saw the practitioner time commitment as reasonable, although some SLT members were unsure of the exact time requirements and based their assessment on conversations they had had with staff. A view among a small group of SLT members was that staff taking part in the training would need to be willing to spend a bit more additional

[20] Exact figures have been suppressed to reduce the risk of disclosure.
[21] The costs estimated by ELRS are based on one form entry schools with one reception class teacher and one TA. The assumption is that handbooks and picture books can be used for the subsequent years so were only included in the estimated costs for the first year.
[22] Exact figures have been suppressed to reduce the risk of disclosure.

time delivering the programme on top of their day job, although this was not explicitly mentioned by any participating teacher or TA.

Evidence of promise

The following section explores questions around evidence of promise, covering research questions 7 to 12. The section focused on the pathways to improvements laid out in the theory of change (Figure 2) with a focus on proximal (short-term) outcomes. As this was a pilot evaluation involving 20 schools taking part in a ten-week programme, quantifying impact at the practitioner or child-level was beyond the scope of the study.

Research question 7. Is there an improved understanding, among TAs and teachers, of how children learn about numbers and operations?

Key findings

The majority of school staff felt positively about the benefits of the programme for practitioner knowledge and understanding. TAs' understanding of how children learn about numbers and operations was mirrored in them better understanding and using key mathematical concepts. Participants identified TAs' exposure to the research evidence as well as ringfenced time to attend training and engage in ongoing support with improved understanding of how children learn numbers and operations.

During and immediately after the programme, the majority of school staff felt positively about the benefits of the programme for practitioner knowledge and understanding. Post-delivery survey respondents were asked whether their own understanding, and for teachers, both their own and their TAs' understanding of how children learn about numbers and operations had improved, with options including 'no, not at all', 'yes, somewhat', 'yes, quite a bit', and 'yes, a lot'. Figure 3 shows that all practitioners reported at least some improvement in understanding immediately after the programme ended, with 22 out of 26 of the respondents reporting that their understanding of how children learn maths had improved 'quite a bit' or 'a lot' post-delivery[23] (Appendix 12 Table 24).

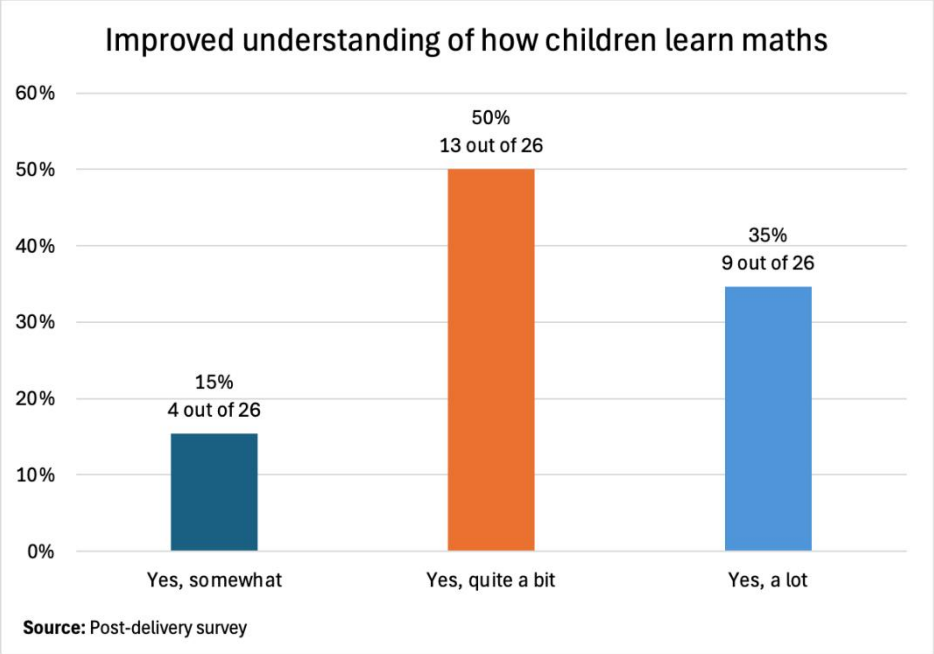


Figure 3: Improvements in practitioner understanding of how children learn maths

[23] Teacher and TA data has not been disaggregated to reduce the risk of disclosure.

If the response options are interpreted as a continuous scale from 0 to 3, with 0 representing 'no, not at all' and 3 representing 'yes, a lot', the results show that teachers were generally more optimistic about the improvement in TAs' understanding than the TAs were themselves. The average teacher-reported TA understanding was 2.375, indicating that teachers believed the TAs' understanding improved between 'quite a bit' and 'a lot' (between 2 and 3). In contrast, self-reported TA understanding had a lower average of 2, suggesting that TAs felt their own improvement was only 'quite a bit' (Appendix 12 Table 25). This may be an indication of a general tendency for respondents to be modest in self-reporting outcomes or overestimate benefits for others. Subgroup analysis conducted on the data revealed no statistically significant difference in reported improvements based on background factors like experience, age, or education level.[24]

Qualitative interviews with teachers, TAs, and SLT members revealed an improvement in TAs' understanding of how children learn about numbers and operations. This was reflected in better knowledge of children's mathematical development and an increased knowledge and use of key concepts.

The programme enabled practitioners to explicitly think about how children learn maths. An SLT member cited the example of the developmental progression, which highlights the paths children follow as they develop their mathematical skills and concepts, to illustrate this:

> *All those sorts of things, I've never explicitly thought about that.* (SLT member)

TAs' understanding of how children learn about numbers and operations was reflected in them better understanding and using key mathematical concepts. The delivery team remarked how TAs displayed confidence and understanding of core fundamental mathematical concepts like cardinality and subitising. Similarly, a teacher explained how their TA applied these key concepts during MTPB sessions:

> *When they're spotting things, she will use words like, oh, we've subitised.* (Teacher)

Participants attributed improvements in TAs' understanding of how children learn numbers and operations to their exposure to the research evidence, supporting the pathway in the theory of change that links engagement with personal development with improved understanding of how children learn numbers and operations. Being exposed to the research evidence enabled them to reflect on their own practice and question entrenched ideas, including how children learn and develop and what 'good teaching' looks like.

Moreover, having time 'ringfenced' to attend training, receive support from teachers, and deliver the intervention sessions enabled TAs' understanding of children's mathematical development to take hold, as outlined in the theory of change. For instance, a teacher explained that they and their TA strictly followed the programme requirements and did not let other priorities get in the way of the key activities.

Research question 8: Is there an increased confidence in teaching maths among TAs and teachers?

Key findings

The majority of teachers and TAs reported positive improvements in confidence after programme completion. Particularly TAs were seen to have improved their confidence in teaching maths and their confidence as practitioners. Participants attributed the increased confidence in teaching maths, particularly among TAs, to their exposure to a targeted CPD programme that was specifically about early years maths.

The majority of teachers and TAs reported positive improvements in confidence after programme completion. Post-delivery survey respondents were asked whether their own confidence, and for teachers, both their own and their TAs' confidence in teaching maths had improved, with options including 'no, not at all', 'yes, somewhat', 'yes, quite a bit', and 'yes, a lot'. Figure 4 shows that a high percentage of both respondents to the post-delivery survey reported improved confidence in teaching maths, with 20 out of 25 respondents saying their confidence has improved 'quite a bit' or 'a lot' post-delivery (Appendix 12 Table 26). Subgroup analysis revealed no statistically significant variation in

[24] Crosstabs have been suppressed due to small cell counts.

improvements in confidence based on background factors.[25] Similar to perceived improvements in understanding (see research question 7 above), teachers' and TAs' views on improvements in TAs' confidence in teaching maths did not align; while teachers reported an average of 2.5 for TAs (indicating an improvement between 'quite a bit' and 'a lot'), TAs themselves reported a lower average of 2.2 (indicating an improvement closer to only 'quite a bit') (Appendix 12 Table 27).

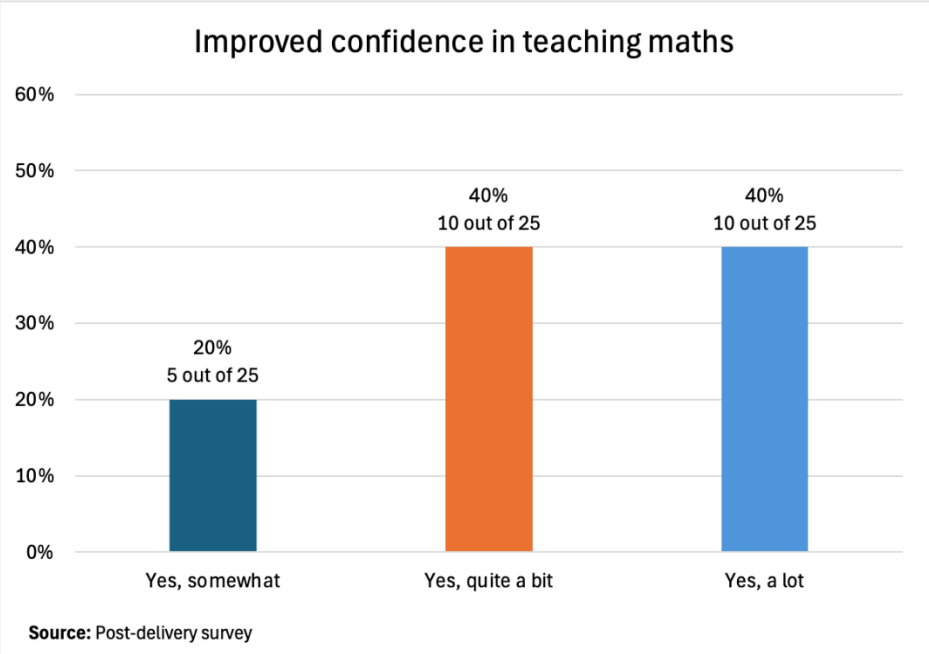


Figure 4: Improved practitioner confidence in teaching maths

Qualitative evidence showed that teachers, TAs, and SLT members felt that practitioners, particularly TAs, improved their confidence in teaching maths and their confidence as practitioners. An SLT member contrasted their teacher' and TA's improvement in maths with the anxiety they felt about maths before the programme, which they believed to be widespread:

> *[Teaching maths] is a huge area of anxiety…I think a lot of teachers and teaching assistants feel very worried about teaching maths.* (SLT member)

Teachers felt that the increased confidence they saw in TAs was reflected in their efforts to incorporate maths into other subjects and daily interactions with children. A TA explained that they had learned new ways to embed maths into literacy teaching, as this also involved reading books. Another TA noted how—following two sessions on the book 'Loo Queue' in Week 3—they now counted when queuing for the toilet:

> *We can get into a queue and go through the numbers.* (TA)

Alongside an improvement in teaching maths, a TA reflected on how the programme helped them become more confident as a practitioner. Specifically, it made them feel more confident delivering sessions in front of a class, which is something that they explained TAs were less familiar with:

> *As a teacher, you'd have to speak in front of a class, but I didn't have that confidence yet. And now I definitely do, even if it's just with a group.* (TA)

Participants attributed the increased confidence in teaching maths, particularly among TAs, to their exposure to a targeted CPD programme that was specifically about early years maths, thus, supporting the pathway laid out in the theory of change.

[25] Crosstabs have been suppressed due to small cell counts.

So, I think that having that focused training has really helped them with their confidence. (SLT member)

The delivery team noted that even those practitioners who had already taken part in maths CPD such as the Maths Mastery enhanced their confidence because the programme helped consolidate core mathematical principles. Participants explained that the programme benefited TAs because teachers often lacked the time to impart relevant subject knowledge to TAs and TAs often lacked confidence with maths. The programme therefore, gave TAs a chance to 'go back to basics' on what and how to teach maths. TAs liked that the programme was specifically targeted at them, which they contrasted with their usual experience of executing sessions planned by the teacher. They found the responsibility of planning and delivering the sessions empowering.

Normally, everything's written up and I'm just delivering what's already been planned. So now it's good because I can say the reasons why and I'm involved in that planning. (TA)

Research question 9. Do pupils talk about number and demonstrate increased enjoyment of and motivation towards learning mathematics?

Key findings

Observations of intervention sessions and interviews highlighted numerous examples of children counting and using mathematical language. Teachers and TAs were convinced that the medium of picture books helped children talk about numbers and carry out mathematical operations. The post-delivery survey indicated that all participants observed at least some improvement in children's enjoyment and motivation towards learning maths. Participants identified the ShREC approach, play-based learning, and group size as key reasons for children's increased enjoyment of and motivations towards learning maths. There was a view among practitioners that children with EAL and SEN particularly benefited from the programme, facilitated by the small group size.

Talking about numbers

Observations of intervention sessions and interviews with teachers, TAs, and SLT members highlighted numerous examples of children counting and using mathematical language. A TA recounted how children made tangible progress throughout the duration of the programme, reflected in their ability to remember numbers from the preceding session as well as adding and subtracting with ease:

[T]hey're really becoming quite confident in their counting and remembering the last number…When I do stop at a number, they do remember that we've stopped at 7 or 9. And even when they are counting when I turn the page, they're able to remember the last number and then add one more and one less. (TA)

TAs also highlighted children's use of mathematical vocabulary as a direct result of TAs modelling language.

From the first to the second session, they've picked up on the language when we say, oh, you've spotted the numeral one, well done. And then now they're using that language. (TA)

Children also demonstrated their understanding of counting non-verbally. A teacher explained how the programme has resulted in children using their fingers to demonstrate numbers in class as well as in MTPB sessions.

If it was number two, one of the little girls [is] holding up two fingers, and she's not saying two, she's just showing it. (TA)

However, observations of intervention sessions showed variation among children. Some children were quicker to demonstrate mathematical knowledge than others. Similarly, some children were less adept at counting on their fingers than others and needed more support from their TA.

There was less evidence of operational vocabulary in observed sessions, with operational mathematical vocabulary often limited to counting. This was especially the case where the observation happened at an early stage of the intervention, when children might not have developed the ability to carry out more complex mathematical tasks. In other cases, though, we observed that the TA did not model operational vocabulary like 'plus' or 'together' for children to replicate.

Teachers and TAs were convinced that the medium of picture books helped children talk about numbers and carry out different mathematical operations, as laid out in the theory of change. They explained that children did not view the sessions as 'learning' or 'doing' maths, which therefore, removed a potential barrier to learning:

> *In their head, we are not doing maths. It's something fun…they're actually learning without realising.* (TA)

> *I don't think they can see the maths learning behind it which is good…it's not like a chore or a task.* (Teacher)

The sequence of the books helped set the expectation that the programme was about maths. Starting with a counting book communicated to children that the programme was about counting—there was a view that this focus on counting could have been lost had the programme started with a book that was less explicitly about numbers.

> *Some of the other books...they obviously do have a mathematical focus, but you could get lost in the story.* (Teacher)

Another group of participants was more reluctant to attribute children's improved counting abilities to the programme alone. For instance, an SLT member noted that their teachers invested a lot of time in improving children's numeracy outside of the programme, which made it more difficult for them to conclude how impactful they believed the programme to be.

Enjoyment and motivation of maths

The post-delivery survey indicated that all participants observed at least some improvement in children's enjoyment and motivation towards learning maths. A majority of respondents—16 out of 26 for enjoyment and 14 out of 25 for motivation—observed improvement in all participating children (Appendix 12 Table 28). However, Figure 5 shows that ten out of the 26 respondents reported that enjoyment improved for only three or four children in their group, indicating that enjoyment did not improve for one or two children in their group. Similarly, 11 out of 25 respondents noted that motivation towards maths improved for only three or four children in their group, indicating that motivation did not improve for one or two children.

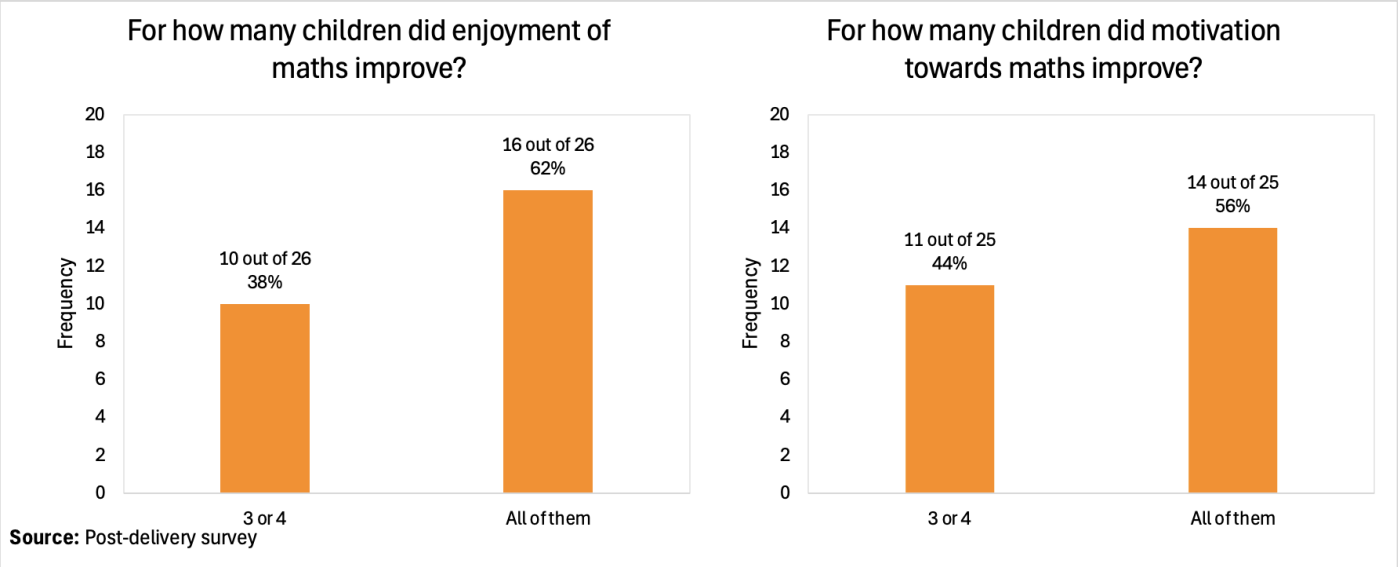


Figure 5: Children's enjoyment and motivation of maths

Observations of intervention sessions and interviews with teachers, TAs, and SLT members demonstrated increased enjoyment and motivation towards learning maths among children, reflected in children's enjoyment of mathematical picture books, and application of learning outside the session. For instance, TAs described how children immediately became engaged as soon as they opened the book:

> *[W]hen I was showing them…maths through a story book…they're getting very excited. And so, like, talking, talking, talking…which is good.* (TA)

Sometimes they just want to immediately start attacking maths as soon as I open the book. (TA)

Children's enjoyment of the books translated into high levels of concentration and focus. A TA pointed this out by describing how none of their pupils had noticed a group of other pupils noisily walking down the corridor, as they had all been fixated on the book:

> *Not one of the children blinked. I [didn't] even know that they'd walked past because we were so engrossed in the book.* (TA)

Participants also cited examples of children sharing their learning with classmates that were not involved in the programme to demonstrate their engagement with maths.

> *One of the children that was in [the] intervention started holding [the picture book] up and modelling what the LSA [learning support assistant] had done during her intervention the first week.* (Teacher)

As with children counting numbers, engagement levels varied among children; session observation showed examples of some children being more active in identifying numbers and objects than others. This difference could have been due to varying levels of interest, confidence, understanding, or familiarity with the content. For instance, during an observation, a child who was particularly engaged explained that they had previously read the book in nursery and so was keen to demonstrate their knowledge of the book to their peers.

Participants identified several reasons for an increase in children's engagement and motivation to learn maths. These related to the ShREC approach, play-based learning, and group size.

Interviews with practitioners highlighted that the move away from asking questions to discussions being child-led increased children's engagement. A TA recounted the example of a child who had become motivated to count and learn maths since taking part in the programme:

> *Before, one child was asked a single question and gave up immediately and said my brain hurts...[Now] they have taken to MTPB.* (TA)

Participants highlighted that the play-based nature of the programme also played a key role in children's engagement. By this they meant the TA's delivery being grounded in humour; the use of picture books rather than numbers to learn maths; as well as the TA's openness to accepting mistakes. This had the cumulative effect that children made progress and learned without realising they were learning.

Learning in a small group format was another factor that drove high engagement levels (see research question 4 above) as laid out in the theory of change. Participants believed a small group environment communicated to children that their voices were being heard and therefore, encouraged participation. Moreover, a small group made it easier for all children to be near the books, which helped with engagement.

> *The books in front of me and I can, you know, I can touch it. They're close, they're involved. I think that's helped.* (Teacher)

There was a view among practitioners that children with EAL and SEN particularly benefited from the programme, facilitated by the small group size.

Post-delivery survey results showed that practitioners felt that certain groups of children experienced different levels of benefit from the MTPB programme. A majority (18 out of 26 respondents) felt that children with EAL benefited more from the programme. Similarly, more respondents reported that children with SEN benefited more from the programme than reporting that they benefited less[26] (Appendix 12 Table 29). During an interview, a teacher described how the programme benefited certain groups of pupils, including children with speech and language difficulties and children with EAL.

[26] Exact figures have been suppressed to reduce the risk of disclosure.

For children with EAL, there was one session that I observed, and that was the most I've ever heard that child [a specific child in the group] speak in school. (Teacher)

Participants attributed the specific benefits for children with EAL and those with speech and language difficulties to ShREC and the small group size. There was a view that the ShREC approach gave children a longer time window to expand on their words without being interrupted. Relatedly, the small size of the group appealed to these pupils more because it allowed them to participate rather than 'hide' in a larger class setting.

They're probably not the children that in a class situation would always be forthcoming. (Teacher)

Children eligible for FSM were perceived to have experienced relatively balanced outcomes, with almost none of the respondents noting that they benefited more or less than their peers.[27]

Research question 10. Is teaching practice embedded and sustained after ten weeks?

Key findings

While the use of picture books to teach maths was largely sustained among nearly all survey respondents, less than half of all survey respondents reported continuing to teach maths with picture books in a small group setting. Similarly, nearly all longitudinal follow-up survey respondents reported continuing using ShREC strategies several months after programme completion.

While the use of picture books to teach maths was largely sustained among nearly all survey respondents, less than half of all survey respondents reported continuing to teach maths with picture books in a small group setting. At the post-delivery survey following programme completion, nearly all respondents expressed a strong intention to sustain key practices, planning to continue using picture books to teach maths and conduct small group reading sessions. The longitudinal follow-up survey conducted ten weeks later confirmed that nearly all respondents continued using picture books, with everyone who initially intended to do so reporting they had sustained this practice. However, only a notable minority continued conducting small group reading sessions after ten weeks.[28]

Similarly, nearly all longitudinal follow-up survey respondents reported continuing using ShREC strategies several months after programme completion. Most respondents had initially planned to sustain these strategies, and this intention was reflected in the longitudinal follow-up survey, where nearly all respondents confirmed they were still using ShREC strategies after ten weeks.[29]

Research question 11. Are there any unintended consequences of participation in this programme?

Key findings

Practitioners used ShREC strategies outside of the interventions, for instance in other subjects. There was some evidence that the intervention had benefits for children's socio-emotional learning, while there was strong evidence that the programme benefited children's language and communication development. Relationships between practitioners and participating children appear to have become strengthened through the programme. There were no concerns expressed about the intervention taking practitioners away from other activities, and limited evidence of children being taken away from other activities. Children selected to take part in the programme did not appear to be stigmatised or feel like they are missing out. It does not appear that unintended consequences affected disadvantaged pupils either more or less than their peers.

Participants valued ShREC strategies, with some practitioners applying these to other subjects.

[27] Exact figures have been suppressed to reduce the risk of disclosure.
[28] Exact figures have been suppressed to reduce the risk of disclosure.
[29] Exact figures have been suppressed to reduce the risk of disclosure.

Participants valued the ShREC approach. One of the benefits teachers and TAs highlighted about ShREC during interviews was its ability to make TAs aware of the knowledge that children had. This would have been less apparent to them had they delivered the session in a less child-led format.

> *It…highlights a lot of knowledge that you didn't know was in there.* (TA)

They also appreciated the structure that ShREC gave practitioners; even where practitioners used a child-led approach in their day-to-day work, ShREC gave them a structure and a set of strategies to draw on.

> *It is an approach that you're sort of mostly doing all the time without realising. But it just gives you a few more ideas of how to start those conversations.* (Teacher)

These perceived benefits of ShREC led practitioners to use the strategies outside of the interventions. A group of teachers and SLT members described applying ShREC strategies to other subjects or when playing with children.

The delivery team stated that their feedback from practitioners showed an appetite to continue the programme in the following academic year, with teachers and TAs expressing a desire to embed ShREC across their daily practice to improve the quality of teaching. Indeed, one teacher we interviewed wanted as many pupils as possible to benefit from ShREC strategies using mathematical picture books, though they noted that this was not feasible with a large group of children:

> *What we would probably like to do is to bring the books in more to actually our whole teaching, although it's quite a difficult thing obviously to teach a whole class in that way.* (Teacher)

Our findings found some evidence that the intervention had benefits for children's socio-emotional learning, while there was strong evidence that the programme benefited children's language and communication development.

<u>Socio-emotional learning</u>

Improvements in children's socio-emotional learning were perceived by 17 out of 26 survey respondents as a result of the MTPB programme (Appendix 12 Table 30).

Observations of intervention sessions demonstrated opportunities for children's socio-emotional learning, reflected in children listening to one another and demonstrating self-control. For instance, during one observation, children added comments in response to other children's contributions; when one child counted 22 fish, another child wanted to try and count all the fish on the page again. Improved turn-taking among children was also emphasised during qualitative interviews with teachers and TAs.

> *Towards the beginning of the programme, they'd like all jump at each other at once and like they'll jump at the book and…now…they're...waiting. And then they'll wait their turn, which is, which is an improvement.* (TA)

Participants who felt that there were socio-emotional benefits identified the ShREC approach and the small group sizes as key factors for improving children's socio-emotional learning. They regarded the ShREC approach as a useful strategy for teaching more 'dominant' children that turn-taking was 'natural'. The small group size equally conveyed to more 'dominant' children the importance of including their peers, because it made them appreciate the impact they had in shaping their immediate environment.

> *They feel compelled to [include quieter child] because they have a bit of a voice in this environment to try and include everybody.* (TA)

However, there were instances where children struggled to take turns without a TA reminding them. For instance, during an observation, three children became distracted, resulting in the TA motivating them with stickers to refocus. A TA similarly described how one 'chatty' child had showed limited improvements in turn-taking and dominated the group.

Language and communication development

Practitioners were overwhelmingly positive about the language and communication development benefits of the MTPB programme, with all 26 respondents in the post-delivery survey indicating that they observed improved language and communication skills in participating children (Appendix 12 Table 30).

Observations of intervention sessions highlighted many examples of children demonstrating language and communication skills, characterised by children making comments, asking questions, using sustained back and forth conversations, and applying non-verbal cues.

Children frequently commented on their observations and linked them to their personal experiences, such as stating that they also had 'three people in my family'. They also asked each other and their TAs questions about the books' content, such as 'Which one looks like a lion?'. This sometimes took the form of sustained back and forth conversation with children listening attentively to each other; in one example, a child counted 33 spots on the fish and was then encouraged by another child to count to 100. Language and communication development also extended to the use of non-verbal cues, with children holding up their fingers to count or using gestures to demonstrate their understanding, such as knocking when the TA read out a specific word.

Interviews with teachers and TAs provided further evidence of children developing their language and communication through participating in the programme. A teacher illustrated this with pupils' increased engagement during story time. Similarly, a TA contrasted their children's language and communication development at the time of the interview with their starting point at the beginning of the programme, noting how much they had improved.

Participants regarded the ShREC approach, the small group size, and the specific picture books as central to improvements in language and communication development.

The ShREC approach was seen as conducive to language and communication development because it combined being child-led—allowing children to talk freely—with practitioners modelling language correctly.

> *It's letting them do most of the discussion...it's me modelling and they're expanding.* (TA)

A TA reflected that the opportunity for children to 'tell a story' improved their confidence. This in turn encouraged them to speak more and develop their language and communication. They reflected how this confidence had particularly helped one of their quieter pupils.

> *She will literally speak over them now and she'll be loud…before we'd have to say, 'come on, use your big voice' and you'd barely hear anything.* (TA)

Alongside the ShREC approach, participants saw the use of specific picture books as another reason for children developing their language and communication. For instance, a TA noted that the colours and different animals in the picture books made it easy for them to expand on children's vocabulary and for children to make direct links with their home life.

The small group size was another factor participants highlighted to explain improvements in children's language and communication. They described that this format and sense of being listened to helped children feel more comfortable to speak compared to a whole-class setting.

However, another group of teachers who we interviewed was reluctant to attribute observed improvements in language and communication to the programme, noting that reception-aged children would be expected to improve their language and communications skills over a ten-week period in the absence of the intervention.

Relationships between practitioners and participating children appear to have become strengthened through the programme. Post-delivery survey results indicated perceived improvements in child–adult relationships, with 21 out of 26 respondents reporting positive changes (Appendix 12 Table 30).

Observations of intervention sessions offered several examples of positive interactions between TAs and their pupils. TAs regularly provided positive reinforcement and personalised the interaction with children, which children responded to positively. For example, during one observation, the TA appeared to incorporate knowledge about children to establish a connection, such as commenting on a child's experiences with their grandparent. In contrast, relationships between

children and the TA seemed more distant where TAs had established strict rules; for example, in a session the TA instructed children to put a finger on their nose when they wanted to speak, which appeared to create a barrier to engagement with the TA.

Interviews with participants echoed many of the positive instances of child–adult relationships visible during intervention session observations. Participants noted examples of children being more communicative and approachable. A TA reflected on children now approaching them outside of the classroom, while another TA explained how one of their pupils has found it easier to express his emotions to them:

> *We've got one child in the group who really struggles with expressing his emotions and he'll just shut off and no one can get out of him what he wants to say. But since doing these sessions, I can bring him out if he's having one of these moments, and he'll tell me whereas he wouldn't before.* (TA)

Participants identified the child-led nature of the programme and the small group size as key in strengthening child–adult relationships. They identified the less-prescribed format of teaching as particularly suitable to forming bonds with children, because it enabled children to take the lead and shape the conversation.

> *There's space for children to talk about their own experience and interests and whatever comes up is explored ...that in itself builds bonds.* (Teacher)

The small group size provided an environment where TAs could dedicate more time to each individual child and identify their learning gaps. It also allowed them to get to know the children better as individuals with unique characteristics.

> *It surprised me, the children's characters are coming out more because it's a small group.* (Teacher)

There were no concerns expressed about the intervention taking practitioners away from other activities, and limited evidence of children being taken away from other activities. A small proportion (four out of 25) of survey respondents in the post-delivery survey reported children's time away from other activities as a negative consequence (Appendix 12 Table 31). Interviewees identified instances where some children wanted to take part in other activities while the MTPB sessions took place, although this did not appear to be a common issue. In one example, the TA and teacher scheduled the intervention session during 'tidy-up time', which some children enjoyed doing.

Children selected to take part in the programme did not appear to be stigmatised or feel like they are missing out. There was no indication from the quantitative and qualitative research activities with teachers, TAs, SLT members, and delivery staff that the children selected for the intervention felt stigmatised. Although not mentioned explicitly by participants, a possible reason for this is that small group activities and moving between rooms is a fairly common occurrence in reception classes, making it less likely for children to feel singled out.

While there was no indication from the quantitative findings that the children not selected for the intervention felt they were missing out, there was some evidence from qualitative interviews with practitioners that this happened. Examples where interviewees brought this up included children wanting to know why they were not included as well as asking to join the sessions as well. A school responded to this by getting the TA to run extra sessions with other groups of children.

It does not appear that unintended consequences affected disadvantaged pupils either more or less than their peers. There is no indication from the qualitative findings with practitioners that unintended consequences affect disadvantaged pupils either more or less than their peers. There was a view from interviews with practitioners, however, that the small group setting particularly helped children with EAL's language and communication development, as they had more time to express themselves and feel listened to.

Research question 12. How different is the programme to business as usual?

Key findings

Schools in the pilot areas use a variety of approaches to teach maths to reception-aged children, and there was variation in schools' use of picture books to teach and embed numeracy. The majority of teachers and TAs who responded to the post-delivery survey believed that the MTPB programme was different to the way maths was normally taught in reception. One way that participants felt that the MTPB programme was different to their usual approaches was that it was more child-led than traditional, teacher-led methods.

Overview of local maths provision for reception-aged children

Our desk-based review of early years maths provision and interviews with Maths Hubs representatives demonstrated that schools in the pilot areas use a variety of approaches to teach maths to reception-aged children. These approaches include memorisation tools (Key Instant Recall Facts or KIRFS), structured lesson plans (Abacus lesson plans and Collins Busy Ant Maths), educational media (BBC Numberblocks), interactive and collaborative methods (songs, role-playing, and carousel), and daily or weekly routines (Mental Maths 5 a Day). This variety reflects the absence of an overarching framework that prescribes a particular teaching approach for this age group.[30]

The desk-based review and interviews with Maths Hubs representatives confirmed that mastery-based programmes and strategies are commonly used in pilot areas, mirroring the increasing relevance of mastery approaches in maths education in England (Boylan *et al.,* 2018). This included:

- **Maths – No Problem!** A series of textbooks recommended by the Department for Education (DfE expert panel to support teaching for mastery.
- **Mastering Number.** Aims to strengthen understanding and fluency with number facts among children in reception and Years 1 and 2.
- **White Rose Maths.** Aims to teach children mathematical concepts in depth before moving to new concepts. While some schools reported using White Rose Maths in reception, others indicated that it was only used from Year 1 onwards.
- **The Concrete, Pictorial, Abstract (CPA) Method.** A pedagogical approach involving the use of concrete materials and pictorial or representational diagrams to build on a child's existing understanding (Yeo *et al.,* 2019).

To support children identified as being at risk of not developing a secure understanding of number before they move to Key Stage 1, teachers and TAs in some schools noted that they had implemented targeted interventions. These took the form of small group sessions with a teacher or TA to provide additional support where it was needed.

Schools also increasingly draw on technology and apps to facilitate maths teaching. Our desk-based review identified a range of apps that schools use, including:

- **Numbots.** A platform that encourages calculation strategies rather than finger counting to teach children addition and subtraction.
- **MyMaths.** An online maths teaching and homework subscription website.
- **Mangahigh.** A digital maths resource for blended learning.
- **RM Easymaths.** Online maths tuition for pupils aged 4–12.

During interviews, teachers and TAs emphasised the role of visuals as crucial in teaching maths, noting that lessons often began with pictorial representations before the class moves on to numerals. They also underlined efforts to integrate maths into everyday activities, linking classroom and real-world experiences. For instance, they described how they practiced counting during baking activities as an example of ensuring that children see practical applications of maths in their daily lives.

Across the pilot schools there was variation in schools' use of picture books to teach and embed numeracy. For instance, a teacher described their approach as 'book centric', while others highlighted that they had never previously used picture books to teach maths.

CPD provision for reception-based practitioners

[30] While the Early Years Foundation Stage (EYFS) statutory framework provides a clear set of expectations around what children must learn, it does not dictate how this learning should be achieved (Department for Education, 2023b).

Schools exercise discretion in selecting their approach to CPD in the absence of a borough-wide approach.[31] Our desk-based review and interviews with Maths Hubs representatives highlighted Teaching School Hubs and Teaching School Alliances as significant sources of CPD alongside commercial organisations offering subscription packages independently from local authorities. Maths Hubs representatives also noted less formal, collaborative networks with other schools or education institutions, who have partnered together to share knowledge, resources, and best practices. Maths-specific CPD is provided through Maths Hubs primarily, with some resources also being provided through independent education consultants.

Participants discussed differences between MTPB and their usual practice (business as usual) with reference to: i) ShREC; ii) teaching early years maths; and iii) the environment.

Comparing MTPB with business as usual

The majority (17 out of 26) of teachers and TAs who responded to the post-delivery survey believed that MTPB was either 'somewhat different' or 'very different' to the way maths was normally taught in reception, with nine out of 26 seeing it as 'somewhat similar' or 'very similar' (Appendix 12 Table 32).

One way that participants felt that MTPB was different to their usual approaches was that it was more child-led than traditional, teacher-led methods. This created space for children's opinions rather than imposing binary understandings of right and wrong. And it allowed teaching to take place at the level of the child.

> *In a scheme it can be very difficult to meet children where they're at, whereas with [MTPB] you can pull out the maths that is appropriate for the children.* (Teacher)

Because MTPB discouraged practitioners from asking many questions, participants explained how they needed to unlearn entrenched habits.

> *Taking a step back [not asking questions] goes against what we've been promoting for so long, to ask questions and get it out [of] the children.* (Teacher)

For a group of participants, MTPB was less structured and prescribed than their usual approach to teaching maths, such as writing out numbers. Others identified similarities with their usual approach; for instance, an SLT member compared the intervention with White Rose Maths and noted that both programmes had rigid planning requirements.

> *The lesson planned script…was quite rigid but then to be honest we follow White Rose and that has a very similar feel in that that has quite rigid lesson plans. So, it's not terribly different at all.* (SLT member)

During interviews with teachers and TAs, a group of participants noted that the ShREC approach was already familiar to them as it had been used in the classroom setting previously, while another group felt that it was very new.

> *[ShREC is] a completely different thing to what we've experienced before.* (Teacher)

Where participants indicated that they had been applying ShREC, they noted that its use in a small group setting was new, and that using it in this way made it more impactful.

> *I saw more of an effect [from ShREC] because there were fewer children.* (TA)

Some participants identified the use of picture books to teach numeracy as different to their schools' usual approach; it offered fresh opportunities to locate maths learning in areas that had not previously been associated with maths.

Readiness for trial

The following section explores questions around readiness for trial, covering research questions 13 to 15. The section draws on the interview with the delivery team, conversations with the delivery team during the learning workshop, as

[31] The DfE and Office for Standards in Education (Ofsted) do not outline statutory requirements on the content, duration, or delivery of CPD in schools.

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.

Research question 13. Can the programme feasibly be delivered at a larger scale?

Key findings

The programme can be feasibly delivered at a larger scale with some modifications. The delivery team's views of modifications required for scaling delivery relate primarily to the provision of training and support and not intervention sessions. Key modifications include: using mentors to provide ongoing support; increasing the number of reflection sessions supported by TAs videoing themselves; having more structured peer support to generate a community of practice; and only using books that are available in the UK and are written for a UK audience.

The delivery team's plans for what scaling to deliver in more schools would look like is clear and realistic, although there is variation in the level of detail and planning that has gone into each element. Importantly, the intervention's core components can be delivered at a larger scale, this includes: selecting five children who need extra help with early numeracy; delivering biweekly 20-minute intervention sessions in a quiet space; using the correct book each week; not using any resources other than the book; and having a planning meeting between the teacher and TA.

Recruitment of schools

The delivery team have a clearly formulated plan for reaching schools and marketing the programme but have not yet formulated detailed plans about SLT involvement and, which schools would benefit most from the intervention.

Marketing the programme

The delivery team are experienced at marketing interventions and have identified the Maths Hub Network and local authorities as 'levers' they can pull to reach a wide range of schools.

Identifying target schools

The delivery team are still in the process of deciding—with support from the pilot findings—what type of school they would ideally want to target to deliver the programme at scale, namely, what contextual factors schools should possess (because those schools would most benefit from the intervention or be most willing to commit the time and money required). While we did not systematically collect contextual data about schools, interviews with teachers and TAs demonstrated that schools in this pilot varied in terms of the level of disadvantage among pupils (i.e. proportion of pupils eligible for FSM), the number of reception classes, the class size, the provision of quiet space for intervention sessions, and the type of area (e.g. pilot schools included schools in densely populated urban areas and more rural coastal areas). During the learning workshop, the delivery team explained that those schools likely to benefit most are those that lack targeted training for TAs, have limited support from partner schools, or have a high proportion of children struggling with early numeracy. However, such schools are likely to find implementation most challenging, for instance due to the high demands placed on TAs and high levels of staff turnover.

Stipulating SLT involvement

The delivery team still need to clarify the level of SLT involvement. They are clear about the important role SLT members play in facilitating implementation (e.g. ensuring a dedicated space is available and protecting teacher and TA time), which was also partly reflected in the post-delivery survey; for instance, seven out of 25 survey respondents agreed that SLT support made it easier for them to access the training and support provided by the delivery team (Appendix 12 Table 8). The delivery team were conscious that they do not want to dissuade schools from taking part in a trial by having too many requirements for SLT involvement. One option under review is offering short briefings for SLT members on the EEF effective CPD guidance report (Collin and Smith, 2021) before their school completes an MoU, as this could help them understand the mechanisms required to ensure the intervention benefits their staff and pupils.

Pupil selection

Outside of the context of an evaluation, schools select pupils after attending the training. This ensures that teachers have a thorough understanding of the programme and early mathematical development, enabling them to make an informed decision about pupil selection. Although it is possible that selecting pupils pre-training increases the likelihood of teachers selecting the wrong pupils, as teachers will have decided without having attended training, our findings suggest that teachers tended to select the children that needed the additional support with the support of an 'objective' task. In the context of a possible efficacy trial—where pupil selection will need to happen pre-testing and pre-training—we believe the risk of teachers selecting pupils that will not be suitable for the programme to be minimal.

Quality assurance

The delivery team acknowledged that requiring schools to submit planning and reflection templates via Google Form or email is not scalable due to a lack of capacity in the delivery team to analyse the data and the burden on schools. One active consideration is therefore, to 'delegate' the quality assurance function to mentors who would be tasked with supporting, monitoring, and responding to schools' needs. Implementation challenges could be picked up as part of schools submitting delivery logs to mentors and/or during regular reflection meetings with the teacher and TA. Mentors fulfilling the function of quality assuring in-school delivery would also free-up time for the delivery team to quality assure mentor support.

Recruitment of mentors

The delivery team is proposing the use of mentors to support in-school delivery. They are clear about the specific criteria that mentors need to fulfil and are thinking about recruitment channels they would use to recruit them. The delivery team explained that mentors would need to have a good understanding of: i) ShREC; ii) early years pedagogy; and iii) features of high-quality CPD.

Selection of books

During the learning workshop, the delivery team highlighted the importance of using high-quality picture books but raised the risk of being unable to procure enough picture books when delivering at scale. Some of the books were shipped from North America and the delivery team questioned whether they would be able to secure a sufficient number for a trial. The delivery team is reviewing the possibility of mitigating this challenge by providing schools with training and/or guidance on selecting books themselves.

Train-the-trainer

If scaling up for an efficacy trial, the delivery team explained that they would have capacity to deliver the training. However, the delivery team would need to consider a train-the-trainer model to scale up further (see Bury *et al.,* 2024).

The delivery team's views of modifications required for scaling delivery relate primarily to the provision of training and support and not intervention sessions.

The following section includes a range of modifications required for scaling delivery suggested by the delivery team, which are presented along with an accompanying paragraph on the rationale and detail. Importantly, the modifications do not affect the interventions' core components, which are scalable. They build on some of the delivery team's plans of what scaling would look like and relate to support from ELRS, support from teachers, peer support, and selection of books.

Support from the delivery team

Modification 1: Consider having TAs video record their sessions to help reflect on their practice, facilitated by mentors providing ongoing support.

Rationale and detail: Mentors would be well placed to support delivery and behaviour change remotely if the delivery team recruit them based on the selection criteria outlined above. This includes an understanding of ShREC, early years pedagogy, and high-quality CPD. The delivery team highlighted the importance of teachers and TAs both attending reflection sessions to ensure teachers are equipped to support with behaviour change. While videos can be effective

for reflecting on practice and changing behaviour, consideration will need to be given to communicating the value of this to TAs reluctant to watch themselves 'perform'.

Support from teachers

Modification 2: Consider reducing the number of planning sessions and increasing the number of reflection sessions supported by TAs videoing themselves.

Rationale and detail: The delivery team highlighted the need to have more professional conversations to maximise the likelihood of behaviour change. They concluded that three professional conversations throughout delivery were not enough. Initial ideas discussed by the delivery team include TAs videoing themselves at every session and reflecting on their delivery. This would be complemented by regular meetings between the TA, teacher, and a mentor throughout the delivery period.

Peer support

Modification 3: Consider having more structured peer support to generate a community of practice, share best practice, and drive behaviour change.

Rationale and detail: The delivery team acknowledged that relationships and dialogue between participants can be a driver of change. To facilitate this, they suggested the possibility of an online 'hub meeting' where participants would be encouraged to share videos as well as challenges and successes. This would be sandwiched in-between the two in-person training sessions, which currently exist.

Selection of books

Modification 4: Consider only using books that are available in the UK and are written for a UK audience.

Rationale and detail: The delivery team will need to procure significantly more books during a trial and will need to avoid the risk of logistical challenges. Providing guidance to schools on, which books to select rather than prescribing, which books to read might be feasible but carries a high risk of schools selecting books that are not suitable and the programme being delivered with limited fidelity. Moreover, there is a risk schools will be more reluctant to take part if they are tasked with finding resources, as this would be an additional burden on them.

Research question 14. How well do we understand the causal chain and mechanisms which underpin the intended effects of the programme?

Key findings

The causal chain and mechanisms outlining what is expected to lead to change are clearly articulated in the theory of change and draw on existing research evidence.

The causal chain and mechanisms outlining what is expected to lead to change are clearly articulated in the theory of change and draw on existing research evidence. The core components that are essential for the programme to have its desired effect are clearly communicated and largely understood by participants, including: selecting children who need extra help to secure their understanding of numbers and operations; delivering intervention sessions for 20 minutes in a quiet space; using no materials other than the weekly book; and having a planning meeting between the TA and teacher. In particular, the small group format and the quiet space were viewed by participants as beneficial for pupils to become more adept at talking about numbers and operations and higher levels of engagement with numbers.

Research question 15. What potential measures could be used in a subsequent efficacy trial?

Key findings

The analysis showed no significant floor or ceiling effects in either the pre- or post-test distributions, indicating that the outcome measure was appropriately challenging across the range of participant abilities. The WIAT-III UK Maths Problem Solving subtest was generally well-received by teachers and assessors for its age-appropriate and varied

mathematical content, though some items were unfamiliar to reception-aged children. The WIAT-III UK Numeracy subtest was less favourably viewed due to its paper and pencil format and abstract number problems, which did not reflect typical reception classroom practices and caused cognitive overload in children.

Following our desk-based review of outcome measures, we recommended the WIAT-III UK Maths Problem Solving and Numeracy subtests as relevant outcome measures to assess change in children's maths attainment in a potential trial of MTPB (see Appendix 5). Here we present the findings from our analysis of the pilot data, and interviews with teachers and assessors.

Analysis of outcome measure piloting

The outcome measure was piloted with children participating in the programme. The pre-test was conducted in January 2024 (n=97) and the post-test conducted in June 2024 (n=89). Attrition was reasonable. At a retention rate of 92%, it was within the EEF 10% threshold for attrition. Of the 97 children who participated in the pre-test, 89 were assessed at the post-test. Attrition was primarily due to one school dropping out of the pilot.

The analysis showed no significant floor or ceiling effects in either the pre- or post-test distributions. The pre-test scores exhibited a unimodal, slightly skewed distribution, while the post-test scores, which included two subtests, were bimodal. This indicated that the outcome measure was appropriately challenging across the range of participant abilities.

The relationship between pre- and post-test scores was strong, with a Pearson correlation coefficient of 0.688. This statistically significant result reflects consistent measurement across the two timepoints. Additionally, regression analyses confirmed that pre-test performance was a significant predictor of post-test outcomes, with nearly half of the variance in post-test scores explained by pre-test scores.

Intraclass Correlation Coefficients (ICCs) provided insights into clustering within schools. For the primary outcome, the ICC was 0.124, suggesting low to moderate clustering. For secondary outcomes, the ICC was very low for numeracy (0.019) and slightly higher for problem-solving (0.138), implying that school-level clustering should be accounted for primarily in analyses involving problem-solving measures.

A detailed note on the findings of the outcome measure review and piloting can be found in Appendix 5.

Views on outcome measures

The WIAT-III UK Maths Problem Solving subtest was generally well-received by teachers and assessors for its age-appropriate and varied mathematical content, though some items were unfamiliar to reception-aged children. Minor issues with user interface design and scripted guidelines were noted. Teachers appreciated its alignment with the EYFS framework and the Reception Baseline Assessment (RBA). Its interactive iPad format, which engaged children more effectively than paper and pencil assessments, was viewed favourably. Rigid scripting limited assessors' ability to provide minimal prompts, potentially affecting children's performance; however, rigid scripting is necessary for the assessment to be delivered consistently to all pupils.

In contrast, the WIAT-III UK Numeracy subtest was less favourably viewed due to its paper and pencil format and abstract number problems, which did not reflect typical reception classroom practices and caused cognitive overload in children. Challenges included children's unfamiliarity with assessment booklets, answering written sums, high cognitive demands, and assessors' difficulty multitasking between marking and observing. Scripted instructions were perceived as less user-friendly, with assessors noting lack of clarity for handling unexpected scenarios despite guidelines provided.

A detailed note on the views expressed by teachers and assessors on the outcome measures can be found in Appendix 5.

Conclusion

Table 10 below provides a summary of the main findings and Table 11 below includes an overview of our overall assessment of the success indicators. The different success indicators are colour-coded depending on the extent to which we believe they have been met. Where the success indicators have not been fully met, we provide a short explanation about the rationale and an assessment of how much effort will be required to address the challenge.

Table 10: Summary of pilot findings

Research question	Finding
Feasibility Is the approach feasible to implement?	The programme was largely delivered with fidelity. Sessions took place in nearly all schools twice a week during the delivery period and nearly all TAs who completed the survey confirmed the sessions were delivered in a dedicated quiet space. Qualitative interviews with teachers and TAs highlighted that TAs nearly all adopted the ShREC approach, although observations showed that there was some variation in the extent to which they consistently followed it. Challenges to in-school fidelity are mainly related to the ongoing support provided by teachers for TAs. For instance, while planning and reflection sessions happened in nearly all schools, the coverage of the content and quality of support was inconsistent, with some teachers not probing sufficiently or giving little space to their TAs to share their reflections. To address this inconsistency in implementation, the delivery team is considering recruiting mentors to support in-school delivery in the future. Teachers and TAs surveyed were positive about the training and support from the delivery team, and this was mirrored in qualitative interviews with teachers and TAs who highlighted the practical relevance and novelty of the ShREC approach and praised the delivery team's delivery and approachability.
Evidence of promise Is there evidence to support the theory of change?	The quantitative and qualitative findings indicate that the programme has evidence of promise for practitioners and children. The overall majority of teachers and TAs taking part in the survey reported their own understanding of how children learn maths had improved. Similarly, most teachers and TAs who completed the survey reported their own confidence in teaching maths had improved. Qualitative interviews corroborated these findings, with teachers and TAs attributing the increased confidence in teaching maths to their exposure to a targeted and specific CPD programme. The majority of teachers and TAs surveyed, observed improvements in enjoyment and motivation towards learning maths in all participating children. In qualitative interviews, practitioners pointed to an increase in children's enjoyment and motivation as evidenced by high levels of engagement during the sessions. They identified the ShREC approach, play-based learning, and the small group size as key drivers.
Readiness for trial Is the approach ready to be evaluated in a trial?	The programme is ready to be evaluated in a trial. The delivery team and participants highlighted the importance and acceptability of the small group setting, the engaging medium of picture books for teaching numeracy, and the twice weekly 20-minute sessions. The delivery team have a clear and realistic plan for what scaling to deliver in more schools would look like and require and have begun planning for this. Our desk-based research identified an appropriate outcome measure for a possible trial, and pilot data confirmed the measure's effectiveness in capturing the changes we expected to see, showing no floor or ceiling effects at baseline or endline. Feedback from teachers and assessors indicated that the WIAT-III UK Maths Problem Solving used at baseline and endline was generally acceptable and could therefore, be used in a trial. However, the Numeracy subtest was viewed as less suitable for use in schools due to its paper and pencil format and abstract number problems. We therefore, recommend that only the Maths Problem Solving subtest should be considered as an appropriate outcome measure for a trial.

Table 11: Assessment of success indicators

Pilot criteria	Success indicators (research question number)	Evidence of success	Assessment of success
Feasibility of implementation	Programme is delivered with fidelity (research question 1)	• Attendance data shows high levels of in-person training attendance. For instance, all teachers attended the in-person training, and nearly all TAs attended. The in-person training was delivered as intended, with minor adaptations, such as changes in the timings of sessions. Similarly, there were minor adaptations to ongoing support. • Planning sessions happened regularly, with variation in self-reported survey data, however. Most schools submitted the planning templates on a weekly basis based on programmatic data. Qualitative interviews and observations suggested that teachers and TAs mostly but not consistently followed all elements of the planning session. • The reflection sessions between teachers and TAs happened in nearly all schools, although they did not always happen on the day of the observation. Nearly all schools submitted the reflection templates after the reflection sessions were held. As with the planning sessions, qualitative interviews and observations showed that teachers and TAs mostly but not consistently followed all elements of the reflection session. • Intervention sessions took place in nearly all schools twice a week during the delivery period. Nearly all TAs who confirmed the sessions took place in a dedicated quiet space, mirrored in findings from the intervention session observations. Practitioners nearly all adopted the ShREC approach although there was some variation in the extent to which ShREC strategies was consistently followed. • Schools generally did not use any additional resources although observations noted some examples of TAs using props. All observations noted that schools used the correct mathematical picture book.	**Somewhat confident that the challenge can be addressed:** Current challenges to fidelity relate primarily to the ongoing support provided by teachers for TAs and related activities, such as submitting templates. The delivery team are actively thinking about changing the provision of ongoing support for teachers and TAs by pairing teachers and TAs with mentors. The mentors would be tasked with supporting delivery and responding to any emerging issues.
	Pupil selection/eligibility criteria are appropriate and feasible to implement (research question 4)	• Teachers who completed the survey reported following the guidance on pupil selection, with nearly all finding the guidance 'useful' or 'somewhat useful'. Qualitative interviews suggested that schools tended to include some disadvantaged children, children with SEN or EAL; however, the delivery team agreed that a perceived lack of clarity around including pupils with SEN and EAL meant that more schools could have included those groups of children. More than half of all teachers completing the survey believed all children selected were right for the programme. Qualitative data showed that participants who did not think the right children were selected attributed this to children being too 'able' or not all factors were being considered when selecting children, such as considering maths anxiety as a selection criterion. Participants broadly regarded the selection task as straightforward, quick to administer, and relevant.	**Somewhat confident that the challenge can be addressed:** By providing teachers with: i) more information about the children that would benefit from the programme and/or including a list of Frequently Asked Questions to accompany the guidance; and ii) information about the programme, e.g. what interactive reading is about, schools will be able to make a more informed choice.
	Programme inputs (training, physical resources, additional	• All practitioners who responded to the survey found the training and support from the delivery team 'useful' or 'very useful'—this was reflected in the qualitative	**Somewhat confident that the challenge can be addressed:**

	support) are accessible and useful (research question 5)	interviews where participants highlighted the practical relevance and novelty of ShREC and praised the delivery team's delivery and approachability. • Teachers' confidence in carrying out observations and sharing feedback with TAs was more mixed, with half of teachers responding to the survey feeling somewhat confident to do this after their online session. Qualitative interviews similarly highlighted that the online session for teachers could have included more video exemplification; this could have made teachers feel more confident about carrying observations and delivering feedback. • TAs found the support received from their teachers for the planning session useful. There was variation though in the quality of teachers' support: some did not probe sufficiently or took on a dominant role in the planning session. • All the TAs responding to the survey found the reflection session 'useful' or 'very useful'. TAs emphasised in qualitative interviews that these sessions helped them improve their practice. As with planning sessions, there was variation in the feedback given to TAs. • Observations of the intervention sessions showed that the mathematical conversations were pitched at the right level for children. TAs also tended to invite all children to take part. Participants saw the novelty of picture books as particularly suitable and engaging. • Participants felt broadly positive about programme resources provided by the delivery team. They found the reflection and planning templates straightforward and easy to follow, while handbooks served as a useful point of reference throughout delivery.	Teachers' confidence in carrying out observations and sharing feedback with TAs can be addressed by placing more emphasis during the training on behaviour change and professional reflection, as well as ensuring that all training sessions are attended by teachers and TAs. By asking TAs to video each session, TAs will find it easier to identify areas of improvement with support from the teacher and a mentor.
	Programme is affordable to schools (research question 6)	• Nearly all teachers who completed the survey said they did not incur any additional costs during the programme. Qualitative interviews with SLT members highlighted that they generally saw the programme as affordable and believed that trained staff could train others in delivering the programme. SLT members that found the programme to be unaffordable highlighted cover costs as a key barrier. • Most teachers completing the survey reported spending between 15 minutes and 30 minutes a week on average carrying out weekly planning sessions. Self-reported survey data about reflection sessions provided a similar picture. SLT members explained in interviews that they found the time commitment reasonable; however, there was a less common view that participating staff needed to be willing to possibly spend a bit of additional work on top of their day job on programme implementation.	**Somewhat confident that the challenge can be addressed:** It will be important to be fully transparent about the costs that schools might incur as a result of taking part in the programme, so that schools, which are unable to provide internal cover can make a fully informed choice about taking part.
Evidence of promise	Improved understanding among TAs and teachers of how children learn about numbers and operations (research question 7)	• The wide majority of survey respondents (22 out of 26) reported their own understanding of how children learn maths has improved 'quite a bit' or 'a lot' after completing the programme. Qualitative interviews corroborated the quantitative findings, reflected in TAs' understanding and application of key concepts.	**Yes**
	Increased confidence in teaching maths among TAs and teachers (research question 8)	• Most survey respondents reported their own confidence in teaching maths has improved 'quite a bit' or 'a lot'. Qualitative interviews revealed improvements in confidence in teaching maths as evidenced by TAs incorporating maths into other subjects and day-to-day activities with children.	**Yes**

	Children talk about number and operations during the intervention (research question 9)	• Observations of intervention sessions and interviews highlighted numerous examples of children counting (verbally and non-verbally) and using mathematical language. However, there was variation among children, such as some children requiring more support than others.	**Yes**
	Improved enjoyment of and motivation towards learning maths among children (research question 9)	• The majority of survey respondents observed improvements in all participating children (16 out of 26 for enjoyment and 14 out of 25 for motivation). Qualitative findings demonstrated similar findings; participants cited concrete examples of children engaging during sessions and sharing maths learning with their peers outside sessions, although there were also examples of children engaging less with the content and activities.	**Somewhat confident that the challenge can be addressed**: While there is evidence that children enjoyed the sessions, improved enjoyment and motivations towards learning maths did not apply to all children. More specific guidance around the children likely to benefit from the programme might be one way of addressing this.
	Teaching practice embedded and sustained after ten weeks (research question 10)	• A large majority of survey respondents reported they have continued using the ShREC approach and picture books to teach maths with some conducting small group sessions.	**Yes**
	No indication that the programme likely widens the gap between disadvantaged pupils and their peers (research question 11)	• There was no evidence from the survey and the qualitative interviews that the programme widens the gap between disadvantaged pupils, including socio-economically disadvantaged pupils, and their peers. Survey findings, however, suggested the programme may be particularly beneficial for children with EAL and to a lesser extent children with SEN. Qualitative findings supported this, with teachers and TAs suggesting that ShREC and the small group size were particularly suitable to these groups of children.	**Yes**
Readiness for trial	Core programme elements identified (research question 13)	• The core components (key ingredients) are clearly laid out in the theory of change, with evidence collected against the theory of change as part of this evaluation. The core components are clearly articulated to programme participants, e.g. the training slides specify, which programme components constitute 'key ingredients' and are therefore, non-negotiable. Crucially, core components need to be clearly understood for the programme to be scalable. Minor tweaks to the core components of delivery might be required should changes be made to how teachers and TAs will be supported throughout delivery when the programme is delivered at a larger scale.	**Yes**
	Clear and realistic plan for what scaling to deliver in more schools would look like and require (research question 13)	• The delivery team's plans for what scaling to deliver in more schools would look like and require, pertains to: recruitment of mentors; recruitment of schools; quality assurance of delivery; and the procurement of picture books. While there is variation in the level of detail and planning that has gone into each element, the delivery team are clearly engaging with the changes required to reach more schools.	**Somewhat confident that the challenge can be addressed but will require careful and detailed planning:** Recruiting enough suitable mentors to support schools will be challenging, although the delivery team will be able to tap into existing networks to identify suitable mentors. To ensure the programme is impactful and the

			risk of attrition minimised, the delivery team will also need to think carefully about what type of schools to target for a trial. For instance, the need for a small, dedicated space might make the programme more suitable for larger schools.
	Plan for scaling would entail little or no modification to in-school delivery (research question 13)	• There will be no modifications to delivering the MTPB sessions to children. The delivery team and participants highlighted the importance and acceptability of the small group setting, the engaging medium of picture books for teaching numeracy, and the twice weekly 20-minute sessions. • The provision of ongoing support for teachers and TAs—provided remotely by mentors at regular fixed timepoints—would constitute a significant change to in-school delivery aimed at teachers and TAs. It would shift the provision of ongoing support away from the teacher to a mentor and would see the emphasis being placed on professional reflection rather than planning. This proposed change from the delivery team is in direct response to evaluation findings pointing to the variation in quality of teacher support and the value of behaviour change through professional conversations as a key mechanism for impactful delivery.	**Somewhat confident that the challenge can be addressed but will require careful and detailed planning:** The main change to in-school delivery will be the addition of mentors providing ongoing support to teachers and TAs—the main challenge hereby will be ensuring the delivery team will identify the right number of suitable mentors, i.e. mentors familiar with the ShREC approach, early years pedagogy, and the key mechanisms of CPD.
	Causal chain and underpinning mechanisms articulated (research question 14)	• The causal chain and the underpinning mechanisms, i.e. what is expected to lead to change, are clearly articulated in the theory of change, providing a framework for collecting evidence. We have collected promising evidence exploring the causal chain and underpinning mechanisms. There might still be value in revisiting the causal pathway prior to the programme being scaled further.	**Yes**
	Outcome measure(s) identified that is/are: i) aligned to intended outcomes, ii) able to measure changes that we expect to see; and iii) acceptable to schools (research question 15)	• Our desk-based review ensured that the outcome measure was specifically designed to assess changes in mathematics attainment, aligning with the intended outcomes outlined in the theory of change. Pilot data confirmed the measure's effectiveness in capturing the changes we expect to see as a result of the intervention, showing no floor or ceiling effects at baseline or endline. Correlation and regression analyses demonstrated a strong relationship between baseline and endline scores, indicating consistency in pre- and post-test measurement, yet sufficient variation to detect changes in attainment. Feedback from teachers and assessors indicated that the WIAT-III UK Maths Problem Solving subtest was generally acceptable, despite some item-specific challenges, whereas the Numeracy subtest was viewed as less suitable for use in schools.	**Yes**

Formative findings

Across the interviews, teachers, TAs, SLT members, as well as the delivery team made a range of recommendations for refinements and improvements, all of which related to pupil selection and training. Combined with our own reflections, these are presented below.

Pupil selection

Recommendation 1: Consider providing more guidance to programme participants prior to selection: i) about the children likely to benefit from the programme; ii) the extent to which group dynamics/behaviour should be considered when selecting children; and iii) about the programme, for instance what a quiet space is or what interactive reading entails.

Rationale and detail: More guidance on child selection would increase the likelihood of teachers making a more informed choice when selecting children for the intervention and possibly reduce some of the challenges TAs experienced with behaviour management. Providing more information about the programme will better manage expectations and minimise any misunderstandings about the programme.

Recommendation 2: Consider standardising the approach to pupil selection.

Rationale and detail: A more standardised process would ensure that schools follow the same approach to selecting pupils and reduce some of the variation in schools' experiences of the process. One option under review for a potential trial is to require teachers to select a longlist (a list of *all* children who they think would benefit) once they have received information about the programme and key change mechanisms before identifying a final list of five children based on the Give-N task. If delivered outside of the constraints of an evaluation, it is feasible for training to be delivered before pupil selection thus, enabling teachers to have a better understanding of the programme.

Training from delivery team

Recommendation 3: Consider making minor modification to the content of the in-person personal development training covering: i) video reflection; and ii) adaptations.

Rationale and detail: Changes in content should be made to reflect changes to support from the delivery team and teachers. This includes covering the process of video reflection and the logistics of videoing sessions. The delivery team might also want to consider spending even more time covering what is and what is not permitted during intervention sessions and why. This could ensure higher levels of fidelity among TAs and—by extension—better outcomes.

Recommendation 4: Consider complementing the two existing in-person events with a third in-person event midway through delivery.

Rationale and detail: Increasing in-person contact time could help foster relationship-building and collective problem-solving among participants and maintain momentum. Participants were highly positive about the opportunity to meet other local schools and network but had limited interaction with other schools during delivery.

Recommendation 5: Consider teachers and TAs attending all sessions together rather than having a teacher-only session or potentially offering sessions for TAs only.

Rationale and detail: This would reduce the reliance on teachers to cascade learning to TAs and enable TAs to better understand key mechanisms of behaviour change. It would also make the intervention more collaborative and reduce existing power asymmetries, as TAs would be equally involved in training sessions. The delivery team acknowledged the risk of schools needing to release two staff members for training if teachers and TAs attended training together but explained that online sessions could be recorded and made available for those unable to attend.

Pupil testing

Recommendation 6: Consider using only the WIAT-III UK Maths Problem Solving subtest for an efficacy trial. It is recommended that the Numeracy subtest not to be used.

Rationale and detail: The Maths Problem Solving subtest was well-received and appears suitable for an efficacy trial due to its relevance and interactive format. In contrast, the Numeracy subtest faced significant challenges in terms of content, format, and administration, particularly in its alignment with reception classroom practices and its potential to accurately assess mathematical abilities of reception children. This makes it less suitable for an efficacy trial. Moreover, the Maths Problem Solving subtest effectively addresses the mathematical concepts targeted by the programme, making it a suitable standalone measure for assessing the intended outcomes.

Recommendation 7: Continue using external, trained assessors to administer assessments to children in an efficacy trial. The assessors will also need to be blind to allocation to eliminate any bias.

Rationale and detail: Although teachers were confident about administering the tests, employing trained external assessors at trial would ensure consistency and minimise potential biases, ensuring reliability of the data.

Interpretation

The pilot explored the extent to which the programme: i) is feasible to implement; ii) has evidence of promise; and iii) is ready to be evaluated in a trial.

Feasibility of the approach

Findings suggest that the intervention was largely feasible for practitioners to implement.

Participants generally found the **guidance** and accompanying task on selecting children useful. However, a group of participants had mixed views about selecting the right children for the programme, for instance, because some children displayed behavioural problems and disrupted lessons. This demonstrates the need for guidance to be clearer about the importance of group dynamics when selecting children.

In-person training attendance was high, and participants valued that the training was in-person and enabled them to meet fellow practitioners. However, some schools highlighted the challenge of sending TAs and teachers to in-person training events, as this incurred significant cover costs. Participants praised the quality of the delivery and relevance of the training content, noting the practical relevance of the evidence-informed strategies underpinning **ShREC** to their day-to-day work. Attendance of the **online sessions** was more mixed, with the optional drop-in sessions having very few attendees. This illustrates the importance for ELRS to carefully consider the timing of training sessions as well as the messaging. For instance, it may be the case that communicating attendance as *optional* discouraged participation.

There was variation between schools regarding the implementation of **in-school support** and partnership between teachers and TAs. The planning session generally took place weekly and the reflection sessions at three designated timepoints. However, during planning sessions, there was variation regarding the detail and content covered when looking ahead at the following session. Similarly, teachers did not consistently cover all relevant sections during reflection sessions or were able to provide feedback on the same day as observing their TA. It is evident that a lack of time constituted a key barrier for participants to engage more thoroughly with the planning and reflection sessions. Despite this, the evaluation found that TAs found the regular interactions with teachers in the form of planning and reflection meetings valuable, supporting the idea that practitioner practice can be embedded through action planning (Collin and Smith, 2021). This is especially relevant in the context of a programme where the onus of delivery is on TAs and many TAs have limited experience of CPD programmes and early years pedagogy (CFE Research, 2024).

Biweekly sessions (twice a week) were nearly delivered in all schools every week, with most schools spending the suggested 20 minutes on each session. Sessions also took place in quiet areas in nearly all schools, highlighting that it is feasible for most schools to facilitate this. TAs generally used ShREC strategies when delivering MTPB sessions. However, some TAs found the need to use ShREC *and* to have mathematical conversations challenging. This illustrates the importance for TAs to be supported throughout this intervention by a teacher who can offer ongoing support (Sharples *et al.*, 2021). The challenges some TAs experienced of having mathematical conversations also underscores the need for early years settings and schools to invest in developing practitioners' understanding of maths and how this can be applied to effective pedagogy (Clark *et al.*, 2021).

Participants felt positively about programme resources provided by the delivery team, noting that they helped with planning and reflection and were accessible. They enjoyed access to Padlet, which gave them opportunity to consolidate their knowledge and read about other relevant areas without being overwhelmed with information. It was evident during

intervention session observations that children found the **nine specially chosen books** highly engaging, although there were some books, which children found more challenging, for instance because they were written for an American audience.

The composition of the pilot schools was diverse, including: local authority-maintained schools and academies; schools rated 'outstanding' and 'good' by Ofsted; large and small schools; schools in urban, rural, and coastal areas; and schools with a high proportion of disadvantaged pupils eligible for FSM. It does not appear that geographical or demographic factors shaped implementation. Key contextual factors affecting implementation—including the provision of a small space and SLT buy-in—did not appear to be specific to certain types of schools. For this reason and despite the small sample size, we believe the findings to be **generalisable to other schools in England**.

Evidence of promise

The theory of change posits that shorter term outcomes for practitioners based on improved understanding will result in improved confidence and practice (Cross, 2009; Dierking and Fox, 2013; Clark *et al.*, 2021). For pupils, practice and enjoyment are expected to result in improved understanding and eventually learning outcomes (Deans for Impact, 2019). Although the evidence is not consistently strong for all causal pathways, the findings largely demonstrate evidence of promise to support the theory of change.

The findings support the potential for the intervention to change practitioners' **understanding of how children learn about numbers and operations, mirrored in the accurate use of core mathematical concepts**. The wide majority of teachers and TAs who completed the survey reported their own understanding of how children learn maths had improved 'quite a bit' or 'a lot'. Qualitative interviews support the quantitative findings and point to practitioners' exposure to research evidence from credible sources and CPD as a driver of better understanding how children learn about numbers and operations (see Collin and Smith, 2021). This underscores the value of practitioners attending training and the delivery team spending time covering the evidence base and core content with participants before in-school delivery.

The findings highlight an increased **confidence in teaching maths** among teachers and TAs. Most teachers and TAs who completed the survey reported their own confidence in teaching maths had improved, while qualitative interviews revealed improvements in confidence in teaching maths as evidenced by TAs incorporating maths into other subjects and day-to-day activities with children. Participants attributed the increased confidence in teaching maths, particularly among TAs, to their exposure to a targeted CPD programme that was specifically about early years maths. This underlines the relative novelty of a programme that is targeted at upskilling TAs in early years numeracy.

Observations of intervention sessions and interviews with practitioners provided numerous examples of children **talking about numbers** and enjoying the sessions. There was a view among practitioners that the enjoyment stemmed partly from children not realising that they were actively learning, finding pleasure in engaging with specifically chosen picture books, and being 'heard' in a small group setting. The post-delivery survey indicated that all respondents observed at least some improvement in children's enjoyment and motivation towards learning maths.

The use of mathematical picture books was largely sustained, although less than half of respondents to the longitudinal follow-up survey reported continuing to use mathematical picture books in small group settings. It is possible that this is a question of capacity, for instance, not having an additional member of staff available to use picture books with a small number of children. It might also be the case that at the time of completing the follow-up survey respondents were working with a different year group and so no longer used this approach in their practice.

Several of the positive unintended consequences in the theory of change materialised. There was evidence of some practitioners using ShREC outside the intervention. Children's socio-emotional learning was seen to improve in addition to their learning and communication development. In particular, the small group size and the emphasis on child-led communication were seen to contribute to improvements in children's language and communication. There is no indication from the qualitative findings that unintended consequences affected disadvantaged pupils more or less than their peers, although a limitation of these findings is that we did not specify during data collection what we meant by disadvantage. There was a view among participants, however, that the small group setting particularly helped children with EAL's language and communication development, as they had more time to express themselves and felt listened to.

Readiness for trial

The intervention is ready to be evaluated for a trial if certain conditions are met, including developing a clear and realistic plan for scaling that includes the availability of a sufficient number of suitable books.

The core components of the programme are clearly laid out and will only require minor modifications for scaling, such as reducing the teachers' time commitment by limiting the number of planning sessions. The key challenges ahead of a trial for the delivery team are structural but are actionable (see 'Formative findings' section above). 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 (20 schools), but the delivery team are confident they can deliver training at this scale. Discussions about recruiting mentors to support in-school delivery suggest they are already planning for this, and their plans appear realistic and achievable.

The financial cost of the intervention is likely to be fairly low based on the delivery team's estimated costs, although this might change if mentors are recruited to support with ongoing delivery; the time commitment for teachers and especially TAs as well as the cover costs will likely be a bigger barrier than the financial cost of the programme itself.

The piloting of the outcome measure revealed that WIAT-III UK is an appropriate measure for testing children participating in the MTPB programme. It is feasible to administer, appropriately targeted, and sensitive to change.

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 practitioner- or child-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.

We experienced challenges recruiting SLT members throughout the pilot; where SLT members had time to take part in an interview, they often displayed limited knowledge of the programme itself, as they were not tasked with overseeing implementation at their school. The 'SLT voice' is therefore, less prominent within our findings.

We had a low response rate to the longitudinal follow-up survey, and we only draw on the longitudinal follow-up survey findings when reporting on sustained practice. These findings should be treated with particular caution. We were also unable to report on whether practitioner outcomes were sustained due to the low response rate to the longitudinal follow-up survey. Moreover, response was relatively low among TAs in particular at the post-delivery timepoint as well as longitudinal follow-up, and it is possible that TAs who were less engaged with the programme did not take part. Due to the low response rates to the surveys and a small sample size, we suppressed findings where cell counts were less than three to prevent potential disclosure of responses. This meant that we were not always able to make comparisons between teachers' and TAs' views. Moreover, our interview sample of participating teachers did not reflect the ethnic make-up of the programme participants. This suggests that it is possible that other perspectives on the programme have not been included in this evaluation, in particular from those individuals who felt less comfortable taking part in research and evaluation activities. The absence of important voices makes it important for us as researchers to reflect on how we can better include all participants in future evaluations.

We were only able to observe a limited number of observation and planning sessions (three each). They were only meant to happen at three timepoints (reflection sessions) or weekly (observation sessions) and were therefore, more difficult to schedule than biweekly intervention sessions. The limited number of observation and planning sessions however, means that our findings and conclusions related to these activities need to be treated with caution.

When asking participants specifically about pupil disadvantage, we did not explicitly specify what we meant by disadvantage. Participants likely had different understandings of disadvantage, including but not limited to socio-economic disadvantage. On reflection, we should have been more explicit about our definition, which would have provided more confidence in findings related to research questions about disadvantage.

Our findings do not include the voices of children. As explained in Appendix 8, it was difficult to gather meaningful data from children due to methodological challenges, the children's age, as well as the nature of the intervention. While practitioners had a good understanding of children's engagement with the programme and perceived benefits, there can still be value in hearing from young children directly in future evaluations of reception-based programmes where there

is a clear rationale for doing and resource available to take an alternative approach, for example, an ethnographic approach to engaging with children.

Future research and publications

We recommend that a future trial of MTPB seeks to measure change in TA knowledge, understanding, and confidence, as a key causal mechanism in the theory of change. This will depend on appropriate outcome measures as well as sufficient responses from TAs. We experienced challenges securing TAs' participation in evaluation activities. Carrying out focus groups with TAs from different schools could be a possible way to encourage more TAs to take part; the opportunity to be with peers could prove effective in overcoming a reluctance about engaging with research. It could also make TAs feel empowered and feel like their voices are valued in shaping the development of the MTPB programme. Evaluators will need to consider being flexible when scheduling focus groups and give careful attention of how best to frame the 'ask' to maximise TA engagement.

Our evaluation found strong evidence of perceived positive unintended consequences for children and, particularly benefits for their socio-emotional learning. A future evaluation of MTPB could consider whether some benefits currently expressed as unintended consequences could or should be included more intentionally in the theory of change, and evaluated as secondary outcomes.

Future research could also explore the delivery organisation's rationale for selecting specific mathematical picture books and deciding on a specific sequence. This thematic focus was beyond the scope of the pilot but could be important in the context of encouraging schools to select picture books outside of the programme.

Finally, future trials of MTPB could explore differential impacts for different groups of children. In particular, school staff perceived particular benefits for pupils with EAL.

There are no planned publications beyond the evaluation report.

References

Baroody, A.J., and Lai, M. (2022) 'The Development and Assessment of Counting-Based Cardinal Number Concepts'. *Educational Studies in Mathematics*, 111: 2, 185–205. https://doi.org/10.1007/s10649-022-10153-5

Boylan, M., Maxwell, B., Wolstenholme, C., Jay, T. and Demack, S. (2018) 'The Mathematics Teacher Exchange and 'Mastery' in England: The Evidence for the Efficacy of Component Practices'. *Education Sciences*, 8: 4, 202. https://doi.org/10.3390/educsci8040202

Brady, L.M. and Graham, B. (2019) '*Social Research With Children and Young People'*. Bristol: Policy Press.

Bury, J, Marshall, L., Scandone, B. and Turner, F. (2024) 'Train-The-Trainer: Guidance and Recommendations For Teams Scaling Up Education Projects and Programmes'. Oxford and London: Oxford MeasureEd and the EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/train_the_trainer.pdf (accessed 15 March 2025).

Casey, B., Erkut, S., Ceder, I. and Young, J.M. (2008) 'Use of a Storytelling Context to Improve Girls' and Boys' Geometry Skills in Kindergarten'. *Journal of Applied Developmental Psychology,* 29: 1, 29–48. https://doi.org/10.1016/j.appdev.2007.10.005

CFE Research. (2024) *Use of Teaching Assistants in Schools: Research Report'*. London: Department for Education. Available at: https://assets.publishing.service.gov.uk/media/66e31878718edd81771316c5/Use_of_teaching_assistants_in_schools_research_report.pdf (accessed 14 March 2025).

Clark, A., Henderson, P. and Gifford, S. (2021) *'Improving Mathematics in the Early Years and Key Stage 1: Guidance Report'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/early-maths/EEF_Maths_EY_KS1_Guidance_Report.pdf?v=1694062549 (accessed 08 August 2024).

Clements D.H. and Sarama, J. (2018) '*Tips for Read-Alouds in Math'*. Denver, CO: Marsico Institute for Early Learning and Literacy. Available at: https://learningtrajectories.org/documents/1582239622565.pdf (accessed 21 January 2025).

Collin, J. and Smith, E. (2021) *'Effective Professional Development: Guidance Report'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/effective-professional-development (accessed 20 January 2025).

Cross, D.I. (2009) 'Alignment, Cohesion, and Change: Examining Mathematics Teachers' Belief Structures and Their Influence on Instructional Practices'. *Journal of Mathematics Teacher Education*, 12: 325–46. https://doi.org/10.1007/s10857-009-9120-5

Data Protection Act 2018, c. 12. Available at: www.legislation.gov.uk/ukpga/2018/12/contents (accessed 23 June 2025).

Deans for Impact. (2019) '*The Science of Early Learning: How Young Children Develop Agency, Numeracy, and Literacy'*. Austin, TX: Deans for Impact. Available at: https://d10a08pz293654.cloudfront.net/production/documents/The_Science_of_Early_Learning-the-Deans-for-Impact_2020-12-31-165124.pdf?v=1609433485 (accessed 13 March 2025).

Department for Education (DfE). (2017) '*Greater Essex Area Review: Final Report'*. London: Department for Education. Available at: https://assets.publishing.service.gov.uk/media/5a82364d40f0b62305b9313b/Greater_Essex_Area_Review_Report_FINAL.pdf (accessed 25 July 2024).

Department for Education (DfE). (2023a) *'Development Matters: Non-Statutory Curriculum Guidance for The Early Years Foundation Stage'*. London: Department for Education. Available at: https://assets.publishing.service.gov.uk/media/64e6002a20ae890014f26cbc/DfE_Development_Matters_Report_Sep2023.pdf (accessed 29 April 2025).

Department for Education (DfE). (2023b) *'Early Years Foundation Stage Statutory (EYFS) Statutory Framework'*. London: Department for Education. Available at: www.gov.uk/government/publications/early-years-foundation-stage-framework--2 (accessed 25 July 2024).

Department for Education (DfE). (2024a) *'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 (accessed 07 February 2025).

Dierking, R.C. and Fox, R.F. (2013) ' "Changing The Way I Teach" Building Teacher Knowledge, Confidence, and Autonomy'. *Journal of Teacher Education*, 64: 2, 129–44. https://doi.org/10.1177/0022487112462893

Education Endowment Foundation (EEF) (2022) *'Implementation and Process Evaluation Guidance for EEF Evaluations'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/EEF-IPE-Guidance-August-2022.pdf?v=1660741017 (accessed 08 August 2024).

Education Endowment Foundation (EEF). (2023) *'Guidance for EEF Pilot Evaluations'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Pilot-guidance-Oct-2023.pdf?v=1721031784 (accessed 22 January 2025).

Education Policy Institute (EPI). (2024) 'Ofsted Ratings – Have Standards Been Improving?' *Accountability, Assessment, & Inspection Blog.* EPI. Available at: https://epi.org.uk/publications-and-research/ofsted-ratings-have-standards-been-improving/ (accessed 07 February 2025).

Essex County Council. (2021) *'Census 2021 – Ethnicity, Language, and Religion'*. Essex Open Data. Available at: https://data.essex.gov.uk/dataset/ex6gx/census-2021-ethnicity-language-and-religion/ (accessed 25 July 2024).

Hassinger-Das, B., Jordan, N.C. and Dyson, N. (2015) 'Reading Stories to Learn Math: Mathematics Vocabulary Instruction for Children With Early Numeracy Difficulties'. *The Elementary School Journal*, 116: 2, 242–64. https://doi.org/10.1086/683986

Higgins, S., Katsipataki, M., Kokotsaki, D., Coleman, R., Major, L.E. and Coe, R. (2016) '*The Sutton Trust–Education Endowment Foundation Teaching and Learning Toolkit'*. London: Education Endowment Foundation. Available at: https://durham-repository.worktribe.com/output/1606338/the-sutton-trust-education-endowment-foundation-teaching-and-learning-toolkit (accessed 18 June 2025).

Hodgen, J., Foster, C., Marks, R. and Brown, M. (2018) *'Improving Mathematics in Key Stages Two and Three: Evidence Review'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/mathematics-in-key-stages-2-and-3 (accessed 08 August 2024).

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) *'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/IPE_Handbook.pdf?v=1732706091 (accessed 16 August 2024).

James, F. (2022, 24 March) 'The EEF's Early Years Specialist Fliss James Explains How Practitioners Can Use the 'ShREC' Approach to Engage Young Pupils in High Quality Interactions'. *EEF Blog: The ShREC Approach – 4 Evidence-Informed Strategies to Promote High Quality Interactions With Young Children.* EEF. Available at: https://educationendowmentfoundation.org.uk/news/eef-blog-the-shrec-approach-4-evidence-informed-strategies-to-promote-high-quality-interactions-with-young-children (accessed 08 August 2024).

Lewis, J. and Maisuria, A. (2023) ' *'Math to 18' in England'.* Research Briefing No. 9780. London: House of Commons Library. Available at: https://researchbriefings.files.parliament.uk/documents/CBP-9780/CBP-9780.pdf (accessed 16 August 2023).

London Borough of Havering. (2023) *'Havering Demographic Profile: Joint Strategic Needs Assessment 2023–2024'.* Romford: London Borough of Havering. Available at:

https://democracy.havering.gov.uk/documents/s72056/JSNA%20Demography%20Chapter%202023%20v0.3A.pdf (accessed 25 July 2024).

London Borough of Redbridge. (2024) *Joint Strategic Needs Assessment*. Available at: https://www.redbridge.gov.uk/media/12788/jsna-2024.pdf (accessed 18 June 2025).

Michie, S., Van Stralen, M.M., and West, R. (2011) 'The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions'. *Implementation Science*, 6: 42, 1–12. https://doi.org/10.1186/1748-5908-6-42

Milburn, T.F., Girolametto, L., Weitzman, E. and Greenberg, J. (2014) 'Enhancing Preschool Educators' Ability to Facilitate Conversations During Shared Book Reading'. *Journal of Early Childhood Literacy*, 14: 1, 105–40. https://doi.org/10.1177/14687984134782

NCETM (2022) *'The Essence of Mathematics Teaching for Mastery'*. NCETM. Available at: www.ncetm.org.uk/teaching-for-mastery/mastery-explained/the-essence-of-mathematics-teaching-for-mastery/ (accessed 16 August 2024).

Newham Council. (2022) *Employment and Health in Newham: Needs Assessment*: London: Newham Council. Available at: www.newham.gov.uk/downloads/file/5109/employment-and-health-jsna-20220912 (accessed 24 July 2024).

Nicoletti, C. and Rabe, B. (2014) *'School Inputs and Skills: Complementarity and Self-Productivity'*. IZA Discussion Paper No. 8693. Bonn: IZA. Available at: https://docs.iza.org/dp8693.pdf (accessed 18 June 2025).

Office for National Statistics (ONS). (2021a) *'How Life Has Changed in Havering*: *Census 2021'.* ONS. Available at: www.ons.gov.uk/visualisations/censusareachanges/E09000016/ (accessed 30 July 2024).

Office for National Statistics (ONS) (2021b) *'How Life Has Changed in Redbridge: Census 2021'.* ONS. Available at: www.ons.gov.uk/visualisations/censusareachanges/E09000026/ (accessed 25 July 2024).

Office for National Statistics (ONS). (2022) *Regional Ethnic Diversity*. GOV.UK. Available at: www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/regional-ethnic-diversity/latest/ (accessed 30 July 2024).

Office for National Statistics (ONS). (2023) *'Population and Migration Statistics Transformation, Newham Case Study*: *2023'.* ONS. Available at: www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/populationandmigrationstatisticstransformationnewhamcasestudy/2023 (accessed 25 July 2024).

Ofsted. (2021) *'Research Review Series: Mathematics'*. Ofsted. Available at: www.gov.uk/government/publications/research-review-series-mathematics/research-review-series-mathematics (accessed 25 July 2024).

Purpura, D.J., Napoli, A.R., Wehrspann, E.A., and Gold, Z.S. (2017) 'Causal Connections Between Mathematical Language and Mathematical Knowledge: A Dialogic Reading Intervention'. *Journal of Research on Educational Effectiveness*, 10: 1, 116–37. https://doi.org/10.1080/19345747.2016.1204639

Ritchie, J., Lewis, J., McNaughton Nicholls, C. and Ormston, R. (eds.). (2013) *'Qualitative Research Practice'*. London: Sage Publications Ltd.

Salingay, N.R.R. and Tan, D.A. (2018) 'Concrete-Pictorial-Abstract Approach on Students' Attitude and Performance in Mathematics'. *International Journal of Scientific and Technology Research*, 7: 5, 90–111. ISSN 2277-8616.

Sharples, J., Webster, R. and Blatchford, P. (2021) *'Making Best Use of Teaching Assistants: Guidance Report'*. London: Education Endowment Foundation. Available at: https://eef.dev/education-evidence/guidance-reports/teaching-assistants (accessed 08 August 2024).

van den Heuvel-Panhuizen, M., Elia, I. and Robitzsch, A. (2016) 'Effects of Reading Picture Books on Kindergartners' Mathematics Performance'. *Educational Psychology*, 36: 2, 323–46. https://doi.org/10.1080/01443410.2014.963029

Wasik, B. (2008) ‘When Fewer Is More: Small Groups in Early Childhood Classrooms’. *Early Childhood Education Journal*, 35: 515–21. https://doi.org/10.1007/s10643-008-0245-4

Yeo, J.B., Choy, B.H., Lim, L.G.P. and Wong, L.F. (2019) ‘Chapter 8: Innovative Pedagogical Practices’. In: Toh, T.L., Kaur, B., and Tay, E.G. (eds.) *Mathematics Education in Singapore*. Singapore: Springer Nature Singapore Pte Ltd. pp. 165–93.

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

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn