Pilot Evaluation of Stepping Stones Programme
Study Plan

Durham University
Menzies, Wang, Qi and Oughton

Education Endowment Foundation

PROJECT TITLE	Pilot evaluation of Stepping Stones programme
DEVELOPER (INSTITUTION)	Northumbria University
EVALUATOR (INSTITUTION)	Durham University
PRINCIPAL INVESTIGATOR(S)	Vic Menzies and Yuqian (Linda) Wang
EVALUATION PLAN AUTHOR(S)	Vic Menzies, Yuqian Wang, Xiaofei Qi, Rachel Oughton
PUPIL AGE RANGE AND KEY STAGE	6-7 years, Year 2, Key Stage 1
NUMBER OF SCHOOLS/ SETTINGS	15
NUMBER OF PUPILS	400 undertaking programme, evaluation recruiting 150
NUMBER OF TEACHERS	15-30

Evaluation plan version history

VERSION	DATE	REASON FOR REVISION
1.0 [*original*]	09/07/2026	N/A

Background

The importance of Metacognition in mathematics learning

Children's mathematical development extends beyond mastering topics and computational skills to also encompass self-regulation and metacognitive abilities (Education Endowment Foundation, 2021). Research consistently demonstrates that metacognitive skills are strong predictors of mathematics performance in primary pupils (Desoete et al., 2019; Ohtani & Hisasaka 2018; Xie, Zeng & Yang, 2021), and that they have a particularly strong effects on problem-solving (Kuzle, 2018). The positive impact of pupils employing metacognitive *strategies* has been demonstrated across competencies in mathematics and mathematical problem solving (Donovan & Bransford, 2005; Hodgen et al, 2018, Muijs & Bokhove, 2020; Mulholland, 2021; Woodward et al, 2012). In a meta-regression looking at the impact of metacognition and self-regulation interventions, age functioned as a significant moderator of intervention effect size in school-age mathematics. Younger children demonstrated larger effect sizes of these programmes than those aged 8 to 15, with the effect size reaching 0.5 at age seven (Hitt, 2023). Metacognitive teaching approaches are rated as high impact and low cost in the EEF teaching and learning toolkit. However, there is limited guidance and professional development opportunity for teachers to support and develop metacognitive skills in the classroom, meaning that these approaches are often not well understood or implemented (Boylan, et al., 2023; Department for Education, 2024; Jossberger et al., 2010; Muijs & Bokhove, 2020). Hence, it is particularly worthwhile to explore intervention programmes related to metacognition to support teaching approaches for this age group.

Understanding metacognition

Metacognition, commonly described as “thinking about thinking”, refers to “thoughts and feelings about one’s own thinking” (Flavell, 1979, p.908). This concept positions pupils’ learning not as a situation of passive involvement but as one of actively controlling their cognitive process (Livingston, 2003). More precisely, metacognition involves critically evaluating and justifying one’s own knowledge (Son, Furlonge & Agarwal, 2020). This process establishes a feedback mechanism that operates both consciously and automatically, across general and domain-specific contexts (Veenman et al., 2006). Metacognitive leaners actively question their knowledge of task, knowledge of self, and knowledge of strategies, assessing the reliability of their reasoning based on the accuracy of judgements about their own learning, and adjusting their approaches if needed (Education Endowment Foundation, 2025). In this study, reasoning and problem solving are defined in accordance with the National Curriculum (Department for Education, 2021). Reasoning is characterised as the ability to “reason mathematically by following a line of enquiry, conjecturing relationships and generalisations, and developing an argument, justification, or proof using mathematical language” (Department for Education, 2021, Aims section). Problem solving refers to the ability to “apply mathematical knowledge to a variety of routine and non-routine problems with increasing sophistication, including breaking down problems into a series of simpler steps and persevering in seeking solutions” (Department for Education, 2021, Aims section).

Metacognition comprises two interconnected components: knowledge of cognition and regulation of cognition. The former includes declarative knowledge (knowing *what*), procedural knowledge (knowing *how*), and conditional knowledge (knowing *when* and *why*) (Paris et al.,

1983; Pintrich et al., 2000). Regulation of cognition refers to skills—how students manage and regulate their learning through planning, monitoring, and evaluating their cognitive processes (Schraw & Moshman, 1995), all of which are essential to metacognitive control of the learning process (Schraw, 2002). They require students to understand themselves, their goals, and effective strategies for how to achieve those goals. As such, development of metacognition lies beyond normal classroom reflective practice (Wang, Brown & Dawson, 2024), incorporating proactive strategy selection, real-time adjustment during learning, and purposeful modification of future approaches based on past performance—creating a complete cycle of self-directed independent learning.

Teaching metacognitive skills

Structured thinking activities have been viewed as specific pedagogical tools suited to improving metacognition in the primary classroom (Branigan & Donaldson, 2020). They point to the importance of teacher-pupil interactions through elaboration and questioning. Effective metacognitive instruction depends on high-quality professional development (PD), as teachers require explicit training to teach metacognitive skills effectively (Chatzipanteli, et al., 2013; Perry et al., 2018; Chen, et al., 2023). Evidence highlights the need for systematic pedagogical approaches and sustained investment in teacher PD to embed metacognitive instruction in mathematics education.

Effective Teacher Professional Development

Professional development (PD) for teachers has been shown to improve pupil outcomes (Sims et al, 2021; Lynch et al., 2019). Sims et al. identified four broad groups of mechanisms that should be incorporated into teacher professional development to make it more effective. Firstly, PD needs to develop teachers' 'insight', developing their professional understanding of subject knowledge, pedagogical knowledge or understanding of how pupils learn which provides the conceptual basis for change. Secondly, the PD also needs to motivate teachers to change practice ('goals for teaching') providing teachers with the belief that the change is worthwhile, that they have the confidence and self-efficacy to change their practice and that the change aligns with teachers' goals and values. Thirdly, teachers need to develop teaching 'techniques' through practice turning their knowledge into action. This can be implemented through instruction and demonstration, modelling of strategies, practice and rehearsal and through feedback on performance. Fourthly, teachers need to develop their habits of practice allowing them to implement the new practice into their normal teaching. Incorporating PD mechanisms designed to facilitate all four areas led to greater effectiveness of PD and incorporated a greater number of mechanisms also increased the effectiveness of PD programmes previously.

The Stepping Stones professional development model was situated within this research and aimed to incorporate a balanced approach including multiple PD mechanisms across all four groups including PD days to impart information, videos to demonstrate practice, observations and feedback to try out the techniques and a fading model of resource development to ensure that the programme and techniques are embedded in usual practice.

Stepping Stones Programme

Stepping Stones is a PD programme developed at Northumbria University that supports teachers in explicitly teaching metacognitive strategies for solving mathematics problems. It incorporates evidence-based practices in teaching meta-cognition (Education Endowment Foundation, 2021) and uses a fading support model, gradually reducing scaffolding in lesson planning for teachers and in problem solving for pupils throughout the programme.

The programme was developed as an innovation project and has been delivered to seven teachers in five schools through EEF's Early-Stage Programme Development work with positive findings (Mulholland, Gray, Counihan & Nichol, 2025). In this project, teachers viewed the programme positively. They valued the quality of the materials and how well the problem-solving questions aligned with the maths National Curriculum. Teachers found the programme to be accessible, feasible to deliver and relevant, with participating teachers delivering all the programme sessions. Most teachers reported increased use of strategies to support metacognition and problem-solving outside of the programme and that they would continue use of the programme approaches beyond the programme. Some teachers indicated interest from colleagues and wider school adoption of the programme suggesting its perceived value beyond the immediate participants. Participating teachers also reported increased knowledge and confidence in using metacognitive strategies as a result of the programme. The materials that teachers produced during collaborative planning sessions showed that they understood the Stepping Stones strategies and could appropriately apply them when planning their own sessions. The fading PD approach with collaborative planning of sessions during PD days was important to the programme delivery, leading to teachers' sense of ownership and buy in to the programme, and collaborating with other teachers. The research team was seen to deepen teachers understanding and use of metacognitive strategies. Time pressure was identified as a challenge for implementation of Stepping Stones especially during the later stages with more time needed in PD sessions for planning and completing materials.

Feedback from teachers from the initial delivery has informed iterative improvements to the programme including extending the duration of the programme, providing longer PD days for the collaborative development of session materials and greater flexibility in the ordering of the sessions so that it can be aligned to the school's curriculum or maths scheme. The Stepping Stones programme is now ready for piloting across a broader range of schools to assess feasibility, implementation and potential impact on practice and outcomes. This pilot evaluation study aims to do this across 15 schools in the North East and to develop and trial potential measures of metacognition suitable for pupils in Year 2 that could be used in a future randomised controlled trial.

Intervention

A detailed description of the Stepping Stones programme is included in the PLANIT table below (Table 1).

Table 1. PLANIT table for Stepping Stones intervention description

Purpose
1. Name of intervention Stepping Stones Programme

2. Challenge the intervention is designed to address

By supporting pupils' early development of problem-solving skills, learners can be equipped for future success in maths (Woodward et al., 2018). In 2023, 70% of Year 2 pupils met the expected standard in the KS1 SATS assessments. Problem-solving and reasoning accounts for more than half of the marks available on these tests. This indicates that many children in this age range may experience challenges in this vital aspect of the maths curriculum.

The benefits of metacognition are well established in the literature (EEF, 2021 ; Perry et al., 2019) and metacognition can specifically help with mathematical problem solving (Hodgen et al, 2018; Muncer et al., 2022; Vula et al., 2017). However, metacognition is often neglected in policy and practice (Perry et al., 2019) and is often poorly understood by teachers which can lead to misconceptions and weak implementation (Muijs & Bokhove, 2020).

Teachers commonly lack knowledge, confidence, and guidance in how to teach metacognitive strategies effectively; explicit modelling and structured metacognitive instruction are used infrequently in classrooms (Mulholland et al., 2025).

Disadvantaged pupils are less likely to use metacognitive strategies independently without explicit teaching (Muijs & Bokhove, 2020) and therefore may benefit more than other students from structured explicit metacognitive instruction in planning and evaluating their thinking.

The Stepping Stones programme aims to address these challenges through scaffolding teachers' use of evidence-informed strategies to promote metacognition in practice.
Prior pilot work with five schools indicates that teachers value the programme's structure, find it feasible, and report increased use of metacognitive strategies beyond programme sessions.

3. Aims and objectives of intervention

The overall aim of the Stepping Stones programme is to enable Year 2 pupils to understand, apply, monitor, and evaluate metacognitive strategies to support their mathematical problem-solving.

Specifically for pupils it aims to build independence, resilience and confidence in solving multi-step mathematical problems. To do this it aims to improve pupils' ability to articulate their thinking processes and to apply taught metacognitive strategies to their maths problem solving. It also aims to provide pupils with a range of effective problem-solving approaches. Disadvantaged pupils often have fewer opportunities to develop metacognitive skills informally and therefore benefit particularly from the explicit, structured teaching of these skills.

For teachers, the programme aims to increase teachers' knowledge of metacognition and how to teach it explicitly through modelling, worked examples and structured reflection. It also aims to improve teachers' confidence in scaffolding metacognitive talk and integrating metacognitive strategies across wider maths teaching.

4. Core components

The Stepping Stones programme consists of a series of structured lessons which are delivered by teachers to Year 2 pupils three days a week over a 15-week period. The programme consists of pupil-facing materials and teacher scripts to use during lessons, and

structured professional development for teachers which scaffolds the teacher's delivery and planning of these lessons.

Fading Professional Development
The PD programme gradually reduces the support (fading) so that teachers increase their understanding and can plan and model metacognitive strategies to support problem-solving in maths. The PD sessions build teacher knowledge of metacognition, model strategies in actionable ways, and provide dedicated time for collaborative planning of later sessions. This structure ensures teachers become increasingly confident and knowledgeable about using metacognitive practice, and become independent in deploying these practices, supporting sustained behaviour change.

Stepping Stones Lessons
The Stepping Stones lessons and materials involve the explicit teaching of metacognitive strategies, particularly through the use of '*Think Alouds'*, '*worked examples'*, and *structured debriefing questions*. These three aspects are embedded in every lesson to help pupils plan, monitor, and evaluate their mathematical thinking. Repeated opportunities to engage in these strategies are expected to build pupil habits of metacognitive thinking leading to increased ownership of using these strategies independently. It is expected that pupils will grow in confidence and ability to apply their own approaches to problem solving and be better able to explain their thinking.

For the 'Think Aloud' strategy, the programme prompts teachers to narrate their reasoning processes aloud to model the strategy, making previously "invisible" metacognitive processes accessible to pupils, who are subsequently encouraged to 'Think Aloud' themselves in peer and whole-class discussion.

Carefully sequenced worked examples across the 15 weeks of lessons initially present complete step-by-step solutions and gradually shift towards faded, incomplete, or erroneous examples. This design is rooted in cognitive load theory, helping pupils to focus on the reasoning rather than the procedural burden of solving problems. Over time, the faded structure encourages learners to assume increased independence in breaking down problems and selecting effective strategies. Use of these examples is also expected to help pupils understand that there are multiple ways to approach problems and that multiple steps will be involved.

The same four structured debriefing questions are used at the end of each lesson to encourage pupils to reflect on metacognitive processes used during the lesson:

1. What did you learn?
2. What maths knowledge and skills did you use today?
3. What was most challenging?
4. What are your top tips for anyone else looking at this problem?

This allows explicit teaching of the 'evaluation' part of the metacognitive process. The consistent use of the same questions every lesson is expected to help pupils to embed these questions in their maths lessons and to prompt their spontaneous use outside of Stepping Stones lessons.

Facilitative components
Visual metaphor
A further aspect of the programme is the visual metaphor of a river crossed by 'stepping stones', which supports pupils' understanding of multi-step reasoning. Steps in worked examples are physically represented as 'stones' placed across the fabric river, reinforcing the idea that problem-solving involves a sequence of deliberate, connected choices.

Collaborative talk in trios

The programme also relies on structured collaborative talk, with pupils working in mixed-attainment trios discussing question prompts that are part of the lesson slides. High-quality talk is used to promote reasoning, articulate thinking, and compare strategies. Collaborative discussion acts as a mechanism for developing metacognitive understanding and supports disadvantaged pupils in particular, who may have fewer opportunities to develop these skills informally.

How this differs from usual practice

Many aspects of the Stepping Stones programme are different to what would usually be implemented in teaching primary school maths in Year 2. Usual practice in would not typically involve explicit instruction in metacognitive strategies. Research indicates that teachers frequently lack confidence or knowledge in teaching metacognition and tend not to use explicit modelling such as 'Think Alouds', despite evidence of their effectiveness (Mulholland et al., 2025). The systematic and structured approach to worked examples and cognitive scaffolding (including fading and error analysis) would also rarely be used in maths lessons in Key Stage 1.

Usual practice problem solving would often focus on pupils reaching the correct solution rather than the Stepping Stones focus on the reasoning processes for solving the problems. It is also unlikely that Year 2 teachers would have regular and explicit lessons for teaching problem solving strategies. Metacognition and problem solving are often neglected in broader maths curriculum and in teacher PD (Ofsted, 2023).

Finally, the professional development model differentiates the programme from many other CPD programmes. In particular, the fading approach which develops teachers' gradual independence in using and planning lessons is unusual in professional development.

Theory of Change

Teacher causal pathway

Teachers first acquire knowledge of metacognition and related teaching strategies through structured PD. They then observe and rehearse these techniques in early programme weeks using fully scripted materials and receiving feedback on their delivery. This supported delivery and repeated practice using the materials builds teacher confidence in using the materials and strategies. As scaffolding fades and teachers rehearse more through the delivery of the lessons, teachers assume increasing responsibility for planning and delivering sessions, deepening their understanding of how to embed metacognitive talk, modelling, and worked examples. This gradual release leads to increased confidence and capability for using the strategies more widely, enabling teachers to integrate these strategies into broader mathematics teaching beyond the intervention sessions.

Pupil causal pathway

Pupils participate in frequent lessons in which teachers model, and children are asked to engage with, metacognitive strategies related to maths problem solving. Through exposure to worked examples, collaborative discussions in small groups, guided prompts, and visual scaffolds, they build the metacognitive habits of planning, monitoring, and evaluating approaches to mathematical problem-solving tasks. Over time, pupils begin applying these strategies independently, demonstrate greater resilience and willingness to attempt challenging tasks, and take increased responsibility for monitoring and articulating their thinking. These behavioural changes are expected to strengthen their mathematical reasoning and, ultimately, improve maths attainment outcomes. The programme expects particularly strong impact for disadvantaged pupils, who benefit from structured opportunities to develop metacognitive skills.

Learners

5. Intervention recipients (children/pupils/learners)

The programme targets pupils aged 6-7 in their second and third term of Year 2. The intervention is delivered at a whole class level. Pupils who have specific special educational needs which mean they are unable to access the sessions may be excluded from the sessions by exception.

Actors

6. Intervention key facilitator(s)

The intervention is delivered by the Year 2 class teacher. They may be supported by a teaching assistant.

7. Intervention instructor(s)

Professional development training days will be provided by Dr Kirstin Mulholland and William Gray at Northumbria University who developed the programme. Both Kirstin and William are experienced primary school teachers and leaders and have used the approaches embedded in this programme in their own classroom teaching. Kirstin has also previously worked with the Education Endowment Foundation to produce guidance, training and resources for maths, metacognition and teacher professional development, several of which have been incorporated into the 'Stepping Stones' programme.

Observations of practice in settings early in the programme will also be conducted by Kirsten, William and wider members of the delivery team all of whom have good knowledge of the programme and are experienced in delivering it in the classroom. No additional training will be provided to this team as they are already knowledgeable and experienced in programme delivery.

8. Intervention developer(s)

The 'Stepping Stones' programme has been designed by Dr Kirstin Mulholland and William Gray, drawing upon their understanding of research evidence relating to metacognition, mathematical problem solving, and teacher professional development. As described above, Kirstin and William will provide training and support to the class teachers who deliver the programme.

9. Professional development for intervention delivery

The professional development design aligns with the mechanisms identified in the EEF's *Effective Professional Development* Guidance Report (modelling, rehearsal, feedback, collaboration, and planning).

Year 2 teachers responsible for delivering the Stepping Stones lessons participate in:

- Three full-day, face-to-face group training sessions which include collaborative planning activities
- One school-based live observation visit (approximately 90 minutes) during programme weeks 2–3 providing in-school coaching and formative feedback;
- Optional one-to-one online check-in meetings (30–40 minutes) for implementation support;

- Optional online group support sessions (up to one per term, maximum three across the delivery period), scheduled in response to teachers' emerging needs.

This blended model ensures opportunities for modelling, practice, feedback, and reflection—mechanisms highlighted by the EEF as central to effective PD.

The three training days are structured to scaffold teachers' progression from initial knowledge building to independent application.

Training Day 1 (November – before programme delivery)
This full-day session focuses on building foundational knowledge of:

- Metacognition and self-regulated learning;
- The metacognitive strategies central to the programme ('Think Aloud', worked examples, structured debriefing);
- The structure, routines, and resources of the *Stepping Stones* programme;
- High-quality mathematical talk and the TOLD principles.

Teachers engage in interactive activities, including walking through an exemplar session to consider classroom implementation. The day aims to motivate teachers by linking evidence to practice and clarifying the programme's rationale reflecting EEF mechanisms relating to building knowledge and establishing shared goals.

Training Day 2 (February/March)
This session revisits key programme concepts, provides time for teachers to reflect on delivery so far, and offers structured opportunities to address challenges or adaptations identified during implementation. A key element of this day is watching and reflecting on the training video which demonstrates Stepping Stones practice. Another major component is collaborative planning, where teachers work in pairs to create a *Stepping Stones* lesson using the partially completed resources for Weeks 8–11. Draft lessons are shared for feedback, then reviewed by the delivery team and disseminated back to teachers for use. This design provides modelling, rehearsal, feedback, and targeted support.

Training Day 3 (March/April)
The final full-day session prepares teachers for the least-scaffolded phase of the programme. It includes:

- A structured debrief focused on how children are responding to the sessions.
- Guidance on sourcing appropriate problems and structuring sessions independently.
- Substantial time for paired collaborative planning using the blank template materials for Weeks 12–15.

Sessions created by teachers are again reviewed by the delivery team before being redistributed. This approach supports the gradual transfer of responsibility and helps embed practices into day-to-day teaching.

Training uses the full suite of programme materials, including:

- The Stepping Stones programme handbook.
- Annotated example session slides.
- Teacher overview documents and scripts.
- Faded and partially completed session templates.
- Video exemplars.
- Planning templates and key questions for 'Think Alouds', worked examples, and debriefs.

Nature

10. Setting

Stepping Stones takes place within state-funded primary schools. Recruited schools need to have a Year 2 class or a mixed age Year 2 class with only one other year group. Schools across the Northeast of England often have mixed age group classes, represented more frequently across rural areas. The inclusion of these classes in the study allows exploration of how feasible the programme is for these classes which will likely have a greater range of maths abilities, and for mixed Year 1/Year 2 classes, children who have not covered all areas of Year 2 maths curriculum and may not be developmentally ready for aspects of the programme. It also may be more difficult for these classes to find the time for Stepping Stones lessons when balancing other curriculum delivery across two year groups.

11. Intervention location

Stepping Stones lessons take place in the Year 2 classroom. Settings will be recruited from the Northeast of England (Darlington, Durham, Gateshead, Hartlepool, Middlesbrough, Newcastle-upon-Tyne, Northumberland, North Tyneside, Redcar and Cleveland, South Tyneside, Stockton and Sunderland) which includes some of the most deprived regions in England and is within reach of the delivery team for the training sessions and observation visits.

Intervention

12. Delivery – Modes and content

The Stepping Stones programme is delivered face-to-face by Year 2 classroom teachers, using structured lesson materials and visual resources. Each lesson lasts 15–20 minutes and is delivered three times per week over a 15-week period, totalling 45 sessions. Lessons are delivered to the whole class, with pupils working primarily in mixed-attainment trios to facilitate structured mathematical and metacognitive talk. Teachers lead delivery, with pupils learning through guided modelling, collaborative discussion about worked examples, and structured reflection.

For each lesson children are divided into mixed ability trios and sit in their small groups around a large piece of blue cloth on the classroom floor which represents indicate a river. Each trio is provided with a mini whiteboard and pen for the lesson. During the session children are encouraged to discuss their ideas with their trio peers.

Each session is structured around a worked example of a maths problem specifically linked to an area of the National Curriculum and with a focus on specific problem-solving techniques. The purpose of the sessions is not to 'solve' a given problem, but to explain and explore how a solution was reached, compare different approaches and strategies, and to consider the benefits of these for learning.

Teachers are encouraged to adopt the metacognitive strategy of the 'Think Aloud' when delivering programme sessions focusing on planning, monitoring and evaluating an approach. This involves narrating thoughts aloud to demonstrate how learners approach tasks or problems to improve accessibility of these strategies. The teacher overview with suggested script supports teachers in using the 'Think Aloud' strategy during the lessons. Using this strategy models how an expert learner approaches a problem and shows pupils how they can do this themselves. Children are also encouraged to 'Think Aloud' during the

sessions to elicit their thinking and to help teachers to better understand what they know and potential misconceptions.

Sessions begin with the presentation of the problem for the worked example which is placed on one side of the 'river'. Teachers are encouraged to read this out or get children to read the problem in their trio and to elicit some responses from children about the maths that might be needed for the problem. Step 1 is then placed on the 'river' as the first 'stepping stone'. Children are asked to think about and discuss in their trios what the problem is asking them to do before 1-2 groups feedback to the class. The other steps for the problem are then placed as stepping stones crossing the river and read aloud to pupils. Pupils are asked to think about and discuss in their trios the steps taken in the example using the prompts specified in the lesson materials. Following the worked example teachers can choose to use the optional extension alternating examples provided in the resources. At the end of the session children are encouraged to reflect on their learning by discussing in their groups (and a few groups sharing with the class) their answers to four questions in their groups:

- What did you learn?
- What maths knowledge and skills did you use today?
- What was most challenging?
- What are your top tips for anyone else looking at this problem?

The problems, prompts, and scripts change for each session depending on the problem, but the reflection questions stay the same for all sessions. Each session also includes a slide about teamwork how to work in their trios.

Teachers follow pre-prepared lesson slides in Weeks 1–7, lessons planned collaboratively by teachers from partially completed templates in weeks 8-11 and lessons planned collaboratively by teachers using blank templates in Weeks 12–15.

13. Delivery – Positioning

Stepping Stones lessons are delivered during the school day, typically during existing Morning Maths or non-core time and should be in addition to the daily mathematics lesson. Stepping Stones lessons are aligned to the national curriculum and provide structured opportunities to revisit mathematical content and practise metacognitive strategies alongside the main Year 2 curriculum. Previous delivery of the programme with Year 2 teachers found no issues with fitting the Stepping Stones lessons into their timetabling.

Lessons are delivered in the spring and summer school term (with the first PD session for teachers taking place late in the autumn term). This is to allow Year 2 teachers to cover some core aspects of the maths curriculum in the first term which may be needed for pupils to engage in the Stepping Stones lessons.

14. Materials

Programme materials are available to teachers through an online Padlet page. Online materials include:

- Stepping Stones programme handbook which includes details about how the programme is structured, the aims of the programme, the resources, professional development, details of specific programme strategies along with links to further information and resources.
- Resources for each of the three PD training days including:

    - PowerPoint slides for each day with links to the videos and resources used in the sessions
    - Video recordings of training day 1
    - Training videos
    - Example materials for the Stepping Stones lessons
- Resources for each of the 45 sessions including:
    - Fully completed materials for weeks 1-7 of lesson slides with printable problems and teacher overview and script.
    - Partially completed lesson slides for weeks 8-11 containing problems for each session and the initial steps which are recommended for these and blank template for teacher overview documents
    - Skeleton template materials for slides and teacher overview document for weeks 12-15

The Teacher overview and script for each lesson sets out any preparation needed, the learning intentions and how to structure the session in terms of timing and suggested script and questions to use. The printable lesson slides include topic, problem solving learning outcomes, preparation instructions to pupils, guidance on teamwork, the problem, annotated steps for solving the problem, an alternating worked example and debriefing questions.

The lessons are designed to explicitly teach 11 problem solving strategies overall with a maximum of 3 being included in each lesson. Lessons follow a fading approach providing more completed worked examples early in the programme to scaffold learning and gradually reducing the support through introducing examples with mistakes and more faded examples as the weeks progress. Key questions and discussion prompts are included within all session resources to encourage high-quality talk.

The blue fabric to be used as a visual metaphor during the programme is also included in the resources.

15. Implementation – behaviours

The programme design promotes the three behaviours outlined in the EEF Implementation Guidance Report—Engage, Unite, and Reflect—through several planned strategies.

Engage
Teachers are engaged early through a full-day training session introducing the evidence base for metacognition, modelling the programme structure, and establishing clear goals for implementation. The professional development covers why metacognition matters, which motivates and encourages buy-in from teachers. School visits from the delivery team during Weeks 2–3 ensure teachers are implementing the lessons and engages teachers with reflection on their delivery offering personalised feedback and supporting early successes. The subsequent two training sessions offered during the programme encourage continued engagement and reflection about programme delivery. The faded PD model also encourages teachers to take on greater engagement and responsibility for delivery of the lessons.

Unite
Collaborative training and within this, collaborative planning is a central mechanism for uniting staff, with teachers working in pairs or trios with other teachers in their school or with teachers from similar schools to co-design Stepping Stones lessons during Training Days 2 and 3. These activities build shared understanding and collective responsibility for high-quality delivery, consistent with EEF recommendations on collaboration. Teachers also

engage with the delivery team during these sessions, during the lesson observation and through other online support provided.

Reflect
The programme incorporates structured opportunities for teachers to reflect on delivery, emerging barriers and facilitators, through reflection and review discussions at training days, the school visit lesson observation and debrief, and through the collaborative refinement of co-planned sessions during the training days. The professional development sessions encourage teachers to evaluate children's responses, discuss adaptations, and consider how metacognitive strategies are transferring into wider practice. This aligns directly with the EEF emphasis on iterative reflection during implementation.

16. Implementation – contextual factors

Supporting successful implementation
Successful implementation relies on schools ensuring protected time for teachers to attend the three face-to-face days as well as teachers being able to schedule and deliver the three Stepping Stones lessons per week during the programme delivery period. Although a significant commitment for schools, the three days of training is necessary to progressively build teachers skill and knowledge in delivery and to enable the faded PD model building teacher independence with delivery. At the first session teachers are provided with fully completed resources to take away and use in the classroom, while at the second session, once teachers have rehearsed the strategies and lessons, they plan the next lessons using partially complete resources. At the final session teachers plan lessons from scratch using blank templates. The three PD sessions allow networking and sharing of experiences across teachers as well as collaborative working on planning sessions to minimise any planning requirements when back at school. Shorter PD sessions were trialled previously, but teacher feedback indicated that longer PD was required to complete the planning, hence the full day sessions. Research indicates that collaboration in person is more effective and efficient than online collaboration (Anh Do & Thomas, 2025) leading to the decision to hold these sessions in person.

Being upfront about the requirement to attend training when recruiting schools to the project is expected to support release of teaching staff to attend. The provision of curriculum-aligned, detailed, ready-to-use materials in the early weeks and allocating collaborative planning time during PD days is designed to encourage buy-in and minimise the workload for teachers implementing the programme. The time expectation for teachers to prepare for delivering Stepping Stones lessons outside of the PD days, is expected to be similar to the planning time for any normal lesson time based on teacher self-report from earlier programme delivery (Mulholland, Gray & James, 2024). The programme consists of simple repeated routines that are consistent from lessons to lesson which should support the quick learning and embedding of the programme.

Engaging Stakeholders
Class teachers are the primary deliverers, with senior leaders facilitating timetable adjustments, releases for training, and resourcing (e.g., printing materials, providing classroom space for the "river"). Class teachers are engaged through the professional development sessions and supported in their delivery. School leaders are engaged early through training communications. Positive teacher feedback to senior leaders about the training and early delivery is expected to promote wider commitment.

Ensuring clarity and preventing clash with existing practice
All programme components including lesson structures, scripts, lesson slides, and visual resources are clearly defined within the handbook and PD sessions. The intervention is designed to complement (not replace) the mathematics curriculum, aligning lesson content with National Curriculum objectives. This helps avoid conflicts with existing schemes of work and ensures clarity of expectations for teachers.

The fading model also supports alignment with school routines, as teachers increasingly adapt materials to suit their class context while retaining fidelity to core metacognitive principles consistent with EEF guidance on balancing fidelity and adaptation.

Tailoring

17. Planned adaptations

- Teachers are able to adapt the order of the lessons to ensure that they have covered the required maths curriculum area for the lesson at the time of the lesson delivery.
- Teachers can also split pupils into pairs instead of trios for the lessons if this is expected to work better for their setting. The group discussion aspect is important, but the programme would still be delivered with fidelity with pairs.
- Teachers are expected to adapt lessons for later lessons to better suit their class. They are encouraged to do this in the collaborative planning sessions at Training Day 2 and 3. Adaptations will be reviewed by the delivery team before resources are provided back to the group for delivery to ensure they comply with the core Stepping Stones principles. Lessons will still need to include the key pedagogical elements of the 'Think Alouds', the worked examples, the group discussion and the review elements of the programme.
- Teachers can adapt the script for the lessons assuming that the core lesson components are still included.
- Teachers can choose not to use the visual metaphor if the classroom does not have space for this but would still be required to discuss the stepping stones metaphor with pupils emphasising that there are multiple steps to solving problems.

18. Unplanned adaptations (to be completed at reporting stage following delivery)

Not applicable at this stage

Theory of Change

The logic model for the programme, developed by the delivery team and updated by the evaluation team after the set-up meetings, is included below in Figure 1 for the programme background, Figure 2 for the teacher level and Figure 3 for the child level. The causal assumptions are set out in Table 2 and the contextual assumptions for the programme are set out in Table 3.

Many pupils in Key Stage 1 experience difficulties with mathematical problem-solving and reasoning (Woodward et al., 2018). Evidence suggests that metacognition, defined as pupils' ability to plan, monitor, and evaluate their learning, is strongly associated with improved attainment (EEF, 2021; Perry et al., 2019) and maths problem solving (Hodgen et al, 2018; Muncer et al., 2022; Vula et al., 2017. However, these strategies are often under-taught or poorly understood by teachers (Perry et al., 2019; Muijs & Bokhove, 2020). Disadvantaged pupils, in particular, may be less likely to develop metacognitive habits without explicit instruction (Muijs & Bokhove, 2020).

The programme provides high-quality professional development, structured materials, and carefully designed classroom routines which aim to strengthen both teachers' metacognitive pedagogy and pupils' capacity to plan, monitor, and evaluate their own mathematical thinking. Through these inputs, Stepping Stones aims to improve pupils' independence, resilience, and attainment in mathematical problem-solving, with anticipated benefits particularly for disadvantaged pupils who may otherwise have fewer opportunities to acquire metacognitive strategies independently.

The programme provides teachers with professional development opportunities across three collaborative face-to-face training days which provide instruction on metacognition, use of worked examples to promote metacognitive thinking, modelling of the 'Think Aloud' technique and strategies to structure pupil reflection through debriefs. The later two PD sessions include collaborative planning time to support teachers with gaining independence in designing Stepping Stones lessons. Teachers are also provided with lesson planning and delivery materials including fully completed lessons plans and slides for the early sessions, part-completed templates for later sessions and blank templates for the final sessions to enable them to develop independence in planning. Implementation of the programme is supported by an in-person school visit early in delivery providing feedback and opportunity to discuss any issues and by reflection on delivery at each of the training days through collaborative discussion.

The programme asks teachers to implement three 15–20-minute Stepping Stones lessons with the full Year 2 class, each week for fifteen weeks, using the provided materials. During the lessons teachers follow the script to discuss worked examples with the children, model the 'Think Aloud' strategy, encourage discussion on metacognition and ask children to reflect on the lesson using structured debrief prompts. Children work in mixed-attainment trios to promote peer discussion during the lessons, and the visual metaphor of a river with stepping stones is used as a visual anchor and reminder to break problems down into different steps.

These inputs are positioned to build teachers' knowledge, confidence, motivation, and technique, drawing on the EEF's PD mechanisms. The programme design also includes a fading model to gradually increase teachers' autonomy and foster long-term embedding of metacognitive strategies beyond the structured sessions. The fading model is explicitly designed to support teacher continued engagement in the project through minimising of teacher workload outside of the professional development sessions. The programme expects teachers to take increased responsibility for the planning of the lessons through the provision of partially completed (PD session 2) and blank lesson templates and planning resources (PD session 3). Teachers use these resources during the second and third PD sessions to plan two lessons collaboratively in pairs with support from the delivery team. Planned lessons are quality assured by the delivery team and then shared with all teachers to use during the programme. In this way the faded model continues to engage teachers by minimising the workload needed for planning all lessons, keeping planning time to scheduled PD sessions and providing support to teachers, while still increasing teacher ability to plan the lessons independently.

Following training and supported implementation, teachers are expected to develop improved understanding of metacognition and its relevance to mathematical problem-solving and to increase the frequency and quality of metacognitive instruction (including using 'Think Aloud', worked examples and structured debriefs) in their maths teaching practice. Over the course of the programme, they are gradually expected to take ownership of planning sessions as scaffolding is withdrawn, allowing them to design their own Stepping Stones lessons and to apply metacognitive strategies more flexibly across their wider maths teaching. Ultimately, engagement in the programme is expected to lead to teachers having improved knowledge of metacognitive pedagogy, improved confidence in modelling thinking processes and facilitating

metacognitive dialogue and to increase teachers' use of metacognitive strategies during maths lessons beyond the Stepping Stone sessions.

Through regular participation in Stepping Stones sessions pupils are expected to engage regularly in metacognitive talk which is supported by structured questioning and worked examples. They are expected to participate in collaborative problem-solving discussions within their mixed attainment trios. They will build familiarity with a range of problem-solving strategies (e.g. visual representations, working backwards, working systematically). They will also engage in repeated cycles of reflection through debrief questions aimed at supporting children to plan, monitor and evaluate their learning.

Through repeated exposure to the Stepping Stones lessons pupils are expected to build habits of metacognitive thinking including self-questioning and monitoring their own progress during problem solving tasks. They are expected to be more willing to share their reasoning and more able to articulate their approaches after taking part in the programme. It is also expected that pupils will show improved independence and resilience with more challenging problem-solving tasks as a result of the programme. In the longer term, these benefits are expected to lead to improved mathematical attainment for pupils, particularly in reasoning and problem-solving domain, greater confidence and more positive attitudes to maths and tackling non-routine problems.

In summary, the Stepping Stones programme aims to strengthen the mechanisms known to support metacognition (modelling, worked examples, structured reflection, high-quality talk) through high-quality professional development and structured classroom routines. By enabling teachers to embed these strategies in their everyday mathematics teaching, and by giving pupils sustained, scaffolded opportunities to practice them, the programme aims to develop pupils' capacity to think metacognitively and solve mathematical problems effectively and independently. This, in turn, is expected to lead to improved mathematical outcomes and increased confidence, particularly for pupils who may not otherwise access these metacognitive learning tools.

EEF Logic Model - Stepping Stones Pilot

PROBLEM

By supporting pupils' early development of problem-solving skills, we can equip learners for future success in maths (IES, 2018). In 2023, 70% of Year 2 pupils met the expected standard in the KS1 SATs assessments. Problem-solving and reasoning accounts for more than half of the marks available on these tests. Whilst these tests became non-statutory in the 2023-24 academic year, this nonetheless illustrates that many children in this age range may experience challenges in this vital aspect of the maths curriculum.

The EEF (2018) 'Improving Maths at Key Stages 2 and 3' guidance report cites developing metacognitive strategies as a key means of improving problem-solving. We therefore believe that explicit teaching of these strategies in a problem-solving and reasoning context ultimately has the potential to increase pupils' attainment outcomes. However, metacognition is neglected in both UK policy and practice (Perry et al, 2019), and is often described as 'hidden in plain sight' as it is not always well-understood by educators (Quigley & Stringer, 2018).

This programme aims to address these challenges through scaffolding teachers' use of evidence-informed strategies to promote metacognition in practice. Disadvantaged pupils are more likely to benefit from this explicit teaching approach.

TARGET POPULATION

This is a whole-class programme which aims to support children in Year 2 to apply mathematical knowledge in context (problem-solving focus), with an emphasis on increasing independence and resilience. Monitoring and evaluation of this programme should particularly consider disadvantaged children (eligible for PP) to consider the impact of the explicit teaching of metacognitive strategies on this pupil group.

OVERALL AIM

Pupils will understand and use metacognitive strategies to support, and reflect upon the success of, their application of mathematical knowledge in context.

Disadvantaged Year 2 pupils will have attitudes, independence and resilience for mathematical problem-solving.

In the longer term, increased metacognition will result in improved attainment outcomes (for example through increased scores on mathematical reasoning tests, such as non-statutory KS1 SATs).

WHY THE PROGRAMME MIGHT BE PARTICULARLY BENEFICIAL FOR FSM PUPILS

Disadvantaged pupils may be less likely to employ metacognitive strategies without explicit teaching (EEF, 2021), yet there remains relatively little guidance for teachers regarding how to support the development of metacognition. This programme not only provides structured teaching for pupils, but also uses a fading model of professional development to help teachers build knowledge and confidence in supporting the development of metacognition.

Figure 1. Logic model (background)

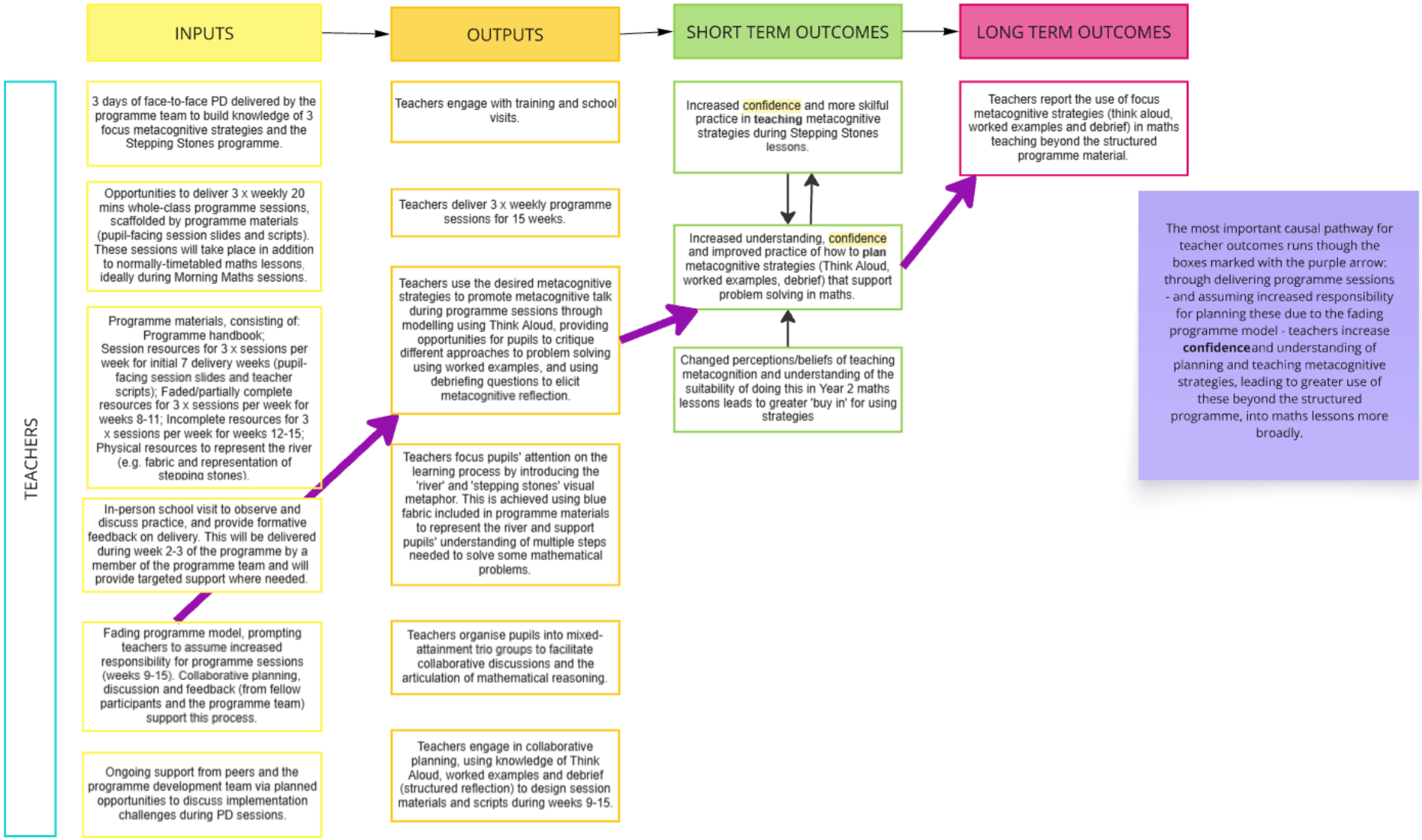

Figure 2. Logic model (teacher level)

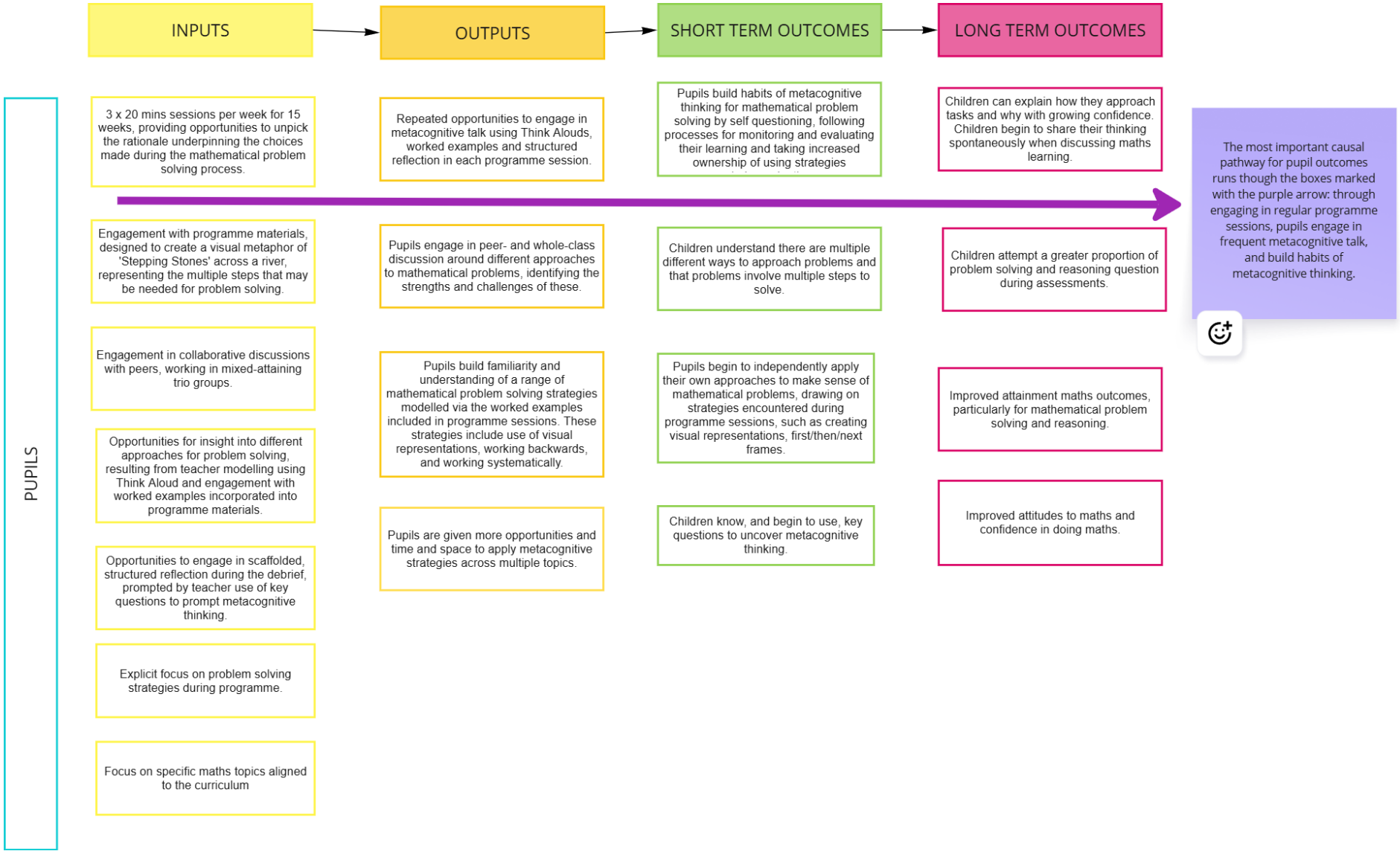

Figure 3. Logic model (Child level)

Table 2. Theory of change causal assumptions

Causal assumption	**Where in the ToC does this assumption occur?**	**Why is this assumption key for your programme?**	**What is the underlying evidence behind this assumption?**	**How strong is the underlying evidence for this assumption?**
Repeated explicit opportunities for both teachers and pupils to engage in metacognitive talk using 'Think Alouds', worked examples and structured reflection (debrief) builds habits of metacognitive thinking and improves maths problem solving, reasoning and overall maths attainment outcomes.	Output, short- and long-term outcomes	This assumption sits at the core of the Stepping Stones programme – if increased engagement in metacognitive talk does not build habits of metacognitive thinking, then long-term outcomes are unlikely to occur. Similarly, if increased engagement in metacognitive thinking does not improve attainment outcomes, then the intended effect of this programme will not be realised.	EEF's Teaching and Learning Toolkit indicates an average impact of +8 months progress in primary school contexts. Dignath, Buettner & Langfeldt's (2008) meta-analysis identified an average effect size of .69 for primary school pupils who had been trained in metacognitive strategies. Perry, Lundie & Golder's (2018) systematic review also concludes that there is 'strong evidence indicating that when metacognition is effectively taught in schools then there is a very positive effect on pupil outcomes'.	**Amber/Green (strong evidence base)** – There are multiple systematic reviews and meta-analyses which indicate the potential value of metacognition for learning outcomes both generally, and for mathematics specifically. However, whilst robust, this evidence is not specific to this programme.
Use of a fading programme model, combined with our professional development (PD) offer, builds confidence, skill and capacity and prompts teachers to assume increased responsibility for programme sessions, supported by collaborative planning and feedback from fellow participants and the programme development team.	Input and output	If PD and programme delivery does not influence long-term behaviour change in terms of encouraging teachers to understand and deploy metacognitive strategies in wider mathematics teaching, then opportunities for promoting the development of metacognition will be missed, with use of these strategies confined to the structured programme sessions.	Sims et al.'s (2021) systematic review commissioned by EEF found an average impact of around +2 months progress for PD programmes incorporating all 14 mechanisms identified. This evidence, and the IGTP model developed from this review, was used to inform our PD model which has been specifically designed to include all 14 identified mechanisms.	**Amber/Green (strong evidence base)** – This evidence is based on findings from the EEF's systematic review of the existing evidence for teacher PD, together with the EEF guidance report derived from this review. It is robust evidence but is not specific to this programme.

Table 3. Theory of change: contextual assumptions

#	**Contextual Assumption**	**Assumption Strength**	**Assumption Risk**
1	Teachers can access the training and will attend the duration, requiring schools to organise cover to release staff.	Amber/Green	Amber/Red
2	Teachers are willing to engage in training.	Green	Amber/Green
3	Teachers currently have gaps in their knowledge of metacognition and teaching strategies associated with this.	Green	Green
4	Teachers currently have gaps in their knowledge of mathematical problem solving and teaching strategies associated with this.	Green	Green
5	Teachers 'buy in' to the importance of metacognition and mathematical problem-solving and are motivated to engage with the training.	Green	Amber/Green
6	Teachers are motivated to engage in collaborative planning and discussion for weeks 8-15 of the programme.	Green	Amber/Green
7	Pupils engaging in the Stepping Stones programme participate fully in sessions, including through discussion in their 'trio' or in pair groups.	Amber/Green	Amber/Red
8	Pupils can access Stepping Stones session content and this is appropriately pitched to match their learning needs.	Green	Amber/Red
9	Leaders 'buy in' to the underpinning pedagogic principles in the Stepping Stones programme (e.g., a commitment to the use of talk to promote learning).	Green	Amber/Red
10	Teachers deploy the Stepping Stones programme as intended, using programme materials and incorporating the focus metacognitive strategies: worked examples, 'Think Aloud', debrief.	Green	Amber/Red
11	Teachers have the capacity to appropriately tailor Stepping Stones sessions content to suit the needs of their cohorts, whilst maintaining fidelity to the underpinning pedagogic principles.	Green	Amber/Green
12	Stepping Stones programme sessions can be timetabled x 3 per week in addition to other maths lessons without major consequences to other learning.	Green	Amber/Green
13	Teachers are given the opportunity through school support to change their practice (e.g., time to prepare Stepping Stones sessions, support from senior leaders).	Amber/Green	Amber/Red

14	The programme development team can access schools to undertake in-school support visits (i.e., schools are within a reasonable distance to enable travel).	Amber/Green	Green
15	Schools will accommodate in-school support visits.	Green	Amber/Red
16	Disadvantaged students can access and engage with the Stepping Stones sessions	Green	Amber
17	Alignment of the programme to the National Curriculum makes it more acceptable for teachers to implement	Green	Green

Research questions

Aims

The evaluation aims to examine whether Stepping Stones shows evidence of promise in achieving the changes articulated across the Theory of Change, with particular focus on the extent to which the programme influences teacher confidence, knowledge and practice. It also aims to assess the feasibility and acceptability of implementation in participating schools with varying school characteristics, considering factors such as training demands, session delivery, and contextual fit. Finally, the evaluation will determine whether the programme is ready to be scaled up for a future trial, identifying any refinements needed to support delivery at larger scale and looking at potential pupil outcome measures.

Evidence of Promise

1. To what extent does Stepping Stones lead to changes in teachers' understanding of and confidence in planning and teaching metacognitive strategies both immediately after the training and through delivery of the programme?
2. How does the classroom practice of teachers taking part in the Stepping Stones programme change as a result of the programme?
3. What is the perceived impact of the Stepping Stones programme on all Year 2 children's ability to solve problems and use metacognitive strategies and for disadvantaged children specifically?
4. Are there any unintended consequences of the Stepping Stones programme?
    a. What have schools stopped doing in order to do Stepping Stones?
    b. How does Stepping Stones fit in within schools' current maths provision?
5. To what extent is the Stepping Stones Theory of Change supported and what are the perceived mechanisms by which the programme works?

Feasibility and acceptability

6. To what extent do teachers engage with the Stepping Stones training and how is this training perceived?
7. To what extent is the Stepping Stones programme being implemented in classrooms as intended?
    a. What adaptations are being made?
8. How acceptable is the Stepping Stones programme to schools?
    a. How do teachers, senior leaders and children (including disadvantaged children) view the Stepping Stones programme?
    b. To what extent are disadvantaged children able to engage with and access the Stepping Stones programme?
9. What challenges or barriers are experiences by teacher in accessing the Stepping Stones training or delivering the Stepping Stones programme?
    a. Are schools able to release staff for Stepping Stones training?
    b. What contextual facilitators specifically support delivery of the Stepping Stones programme in schools?

Readiness for trial

10. To what extent is the Stepping Stones sufficiently manualised to allow consistency in delivery of training sessions, observations and delivery of the programme in schools?
11. Can Stepping Stones feasibly be delivered on a larger scale?
12. What modifications would be required to scale delivery of the programme?
    a. What would be the cost implications of a scaled-up programme for schools? How do schools view the expected scaled-up programme costs?
13. What potential measures could be used in a subsequent efficacy trial to measure metacognitive thinking and problem-solving?

Table 4 below describes the success indicators for each of the three pilot criteria and how they will be assessed during the project. These success criteria were discussed and agreed in the set-up meetings.

Table 4. Pilot Success Criteria

Pilot Criteria	Success Indicators (SI)	How to assess this
Evidence of promise	SI1: At least two thirds of participating Y2 teachers show increased teacher understanding and confidence in teaching metacognitive strategies (RQ1)	Baseline to Endline teacher surveys capturing confidence and knowledge, post-training surveys, interviews with teachers (knowledge of specific terms and understanding). Use of terms in training sessions observations.
	SI2: Changed teacher practice to include metacognition in maths in at least two thirds of participating teachers (RQ2)	Observations of practice using implementation checklist, review of teacher created materials, delivery record in endline surveys, teacher interviews both during the programme and after, baseline to endline teacher surveys.
	SI3: Teachers report that they will continue to teach metacognitive strategies after the programme (RQ2, RQ5)	Endline surveys, post-programme teacher interviews
	SI4: Teachers report that disadvantaged pupils' reasoning and problem-solving ability and willingness to try is improved through the programme (RQ3)	Teacher interviews, endline teacher surveys.

Feasibility and acceptability of implementation	SI5: Programme delivered with fidelity across majority of schools (RQ7) with at least 35/45 sessions delivered by two thirds of schools (with some sessions from all three phases)	Administrative records of school visits with use of fidelity checklists, endline survey, interviews with teachers, developer interview Developer records from training sessions, and emails. Post-training surveys and endline survey to capture sessions delivered.
	SI6: Most teachers attend and engage with all required PD events (minimum one per school) (RQ6)	Administrative records of training attendance, observation of PD events.
	SI7: Majority of pupils (including disadvantaged pupils) engage with the programme content (RQ7, RQ8)	Pupil focus groups, observations of lessons, interviews with teachers.
	SI8: Absence of major barriers (e.g. inability to release teachers for training, inability to fit sessions within current curriculum, inability to fit Stepping Stones sessions into timetable, intervention content perceived as inappropriate for age group/ability) to delivery of the programme in all settings (RQ9)	Interviews with teachers and senior leaders and developer team. Case study observations of practice.
Readiness for trial	SI9: Acceptable plan for modification for delivery at scale maintaining core programme elements and acceptable cost to schools (RQ12 & 13)	Senior leader surveys for acceptable cost. Senior leader and teacher interviews. Workshop and interviews with developer team and EEF review.
	SI10: Sufficient plan for delivery at scale and scaled up programme is similar and maintains core elements of the programme (RQ12 & 13).	Developer interview and workshops. Team judgement.
	SI11: Identification of metacognition assessment options which are aligned to the intervention, appropriate and feasible for teachers to deliver and feasible to deliver at scale. (RQ14)	Recommendation of measures which are acceptable in delivery to schools and provide appropriate data based on piloting and feedback.

Methods

The main research design is a mixed methods implementation and process evaluation including before-and-after intervention quantitative data collection to evidence potential changes, alongside a longitudinal comparative case study design exploring delivery across time and schools. It will also involve cross-sectional surveys following training sessions as well as endpoint and further follow-up interviews with a wider range of schools to support greater generalisation of findings. Triangulation of data sources will strengthen the interpretation of findings.

The project will also feature the development and trial of two maths metacognition assessments suitable for use with Year 2 pupils with the potential to be used in a future randomised controlled trial. A pre-pilot phase will be used to initially develop these assessments working with a separate small group of five schools, while the developed measure will be piloted with more pupils across a further five schools in the pilot evaluation phase.

Recruitment

Pre-pilot assessment development

For the pre-piloting of metacognitive assessments, five schools from across the North East of England will be recruited between March and July 2026. This will allow an iterative process of assessment delivery whereby observation of delivery in each of the five schools informs the trial in the next school making improvements based on findings from the school before. In each of these schools we will work with one Year 2 teacher and/or Teaching Assistant and between four and six Year 2 pupils. These schools will be recruited from the local networks of the project researchers but will aim to recruit a variety of different school characteristics. We will aim to recruit at least one school through delivery team contacts where the setting/teacher has had training or experience with the Stepping Stones programme or has had specific metacognition training. This will help to make sure the assessment is appropriate for schools with and without metacognition expertise as it would be for use at the end of an RCT. Teachers in these schools will be asked to seek parent/carer consent for the child to take part in the assessment trial. Teachers will be asked to recruit pupils will varying maths ability who would be willing to take part in the pre-piloting.

Pre-piloting settings will be given £150 as an incentive for taking part in the project and acknowledge the time required for the work. The Stepping Stones materials will also be made available to these schools. Pre-pilot schools will not be eligible to take part in the pilot evaluation.

Pilot study

The pilot study will work with 15 schools across the North East of England ensuring the inclusion of different school characteristics (e.g. size, location, Ofsted rating and type) to assess how the programme works across different contexts. This number of schools allows the trialling of the programme in a larger number of schools than previously, testing how the PD model scales up. It will also allow for investigation of delivery in a greater variety of school settings, while keeping delivery manageable for the delivery team and costs down. To reach disadvantaged pupils, the majority of schools will be recruited from areas of high socio-economic deprivation aiming for at least five schools recruited with very high levels of deprivation (IDACI deciles 1-3). We will aim to include at least two (and up to three) schools

with mixed age classes (maximum of two year groups per class) to support understanding of how the programme would work in those specific settings where there is likely a higher variation in maths ability and different levels of maths topic coverage. To pilot the programme across settings with different characteristics, no more than two schools would be recruited from the same multi-academy trust. School who are undertaking other metacognition programmes and schools undertaking any other EEF Maths research would be ineligible to participate due to the potential for overlap in programme content and delivery and the burden that undertaking multiple projects could cause for the school.

The developer will lead school recruitment using their networks and contacts with schools across the North East (e.g. Research Schools, Maths Hubs) as well as through advertising through the EEF. Schools will first be provided with an information sheet about the project and using this information would choose to complete an expression of interest form. Schools which complete EOIs will be screened and those eligible will be invited to complete a more in-depth Memorandum of Understanding to sign up to the project. Where more schools are interested than there are places for, additional schools will be placed on a waiting list should any school withdraw before the programme begins.

Across all recruited schools, Year 2 teachers implementing Stepping Stones will be recruited. This must be a minimum of one teacher per school but can be as many teachers as the school can support to attend training.

In five schools selected as case study schools we will also recruit a senior leader and a teaching assistant involved in delivering Stepping Stones lessons (where applicable) for interviews and four to six Year 2 pupils involved in the Stepping Stones lessons in each setting for focus groups. In five non-case study schools, we will recruit up to 10 pupils in each setting to pilot the metacognitive assessment measures. Both case study schools and assessment piloting schools will be given £300 incentive to thank them for their participation in the additional research activities. Case study schools and metacognitive assessment settings will be selected before the Stepping Stones activities begin based on purposive sampling to select schools in both groups with different setting characteristics so that the sample includes at least one larger school, one small school, one school with mixed age classes, include at least three schools with high levels of deprivation (based on FSM percentage), at least one MAT school and schools with different Ofsted ratings. The aim is to select a sample that allows exploration of Stepping Stones practice in a variety of contrasting settings.

Measures and Data Collection

Survey measures

All survey measures will be delivered through the online survey platform Qualtrics with access through a secure Durham University portal. Paper copies of surveys posted to schools may also be used for the endline survey if survey response rates are low – this may be the case as this survey cannot be completed during a training session. Although participants will be asked to complete some of the surveys as part of the training sessions, the information sheet and the verbal instructions given will emphasise that it will be the evaluation team and not the delivery team, who will see and analyse the data and that participants will not be identified in what they say. Any shared findings with the delivery team will be completely anonymised.

Teacher Baseline and Post-intervention surveys

Baseline and post-intervention surveys from practitioners will collect measures of:

- Knowledge of metacognition and teaching metacognition
- Confidence in teaching metacognitive skill in the classroom
- Usual practice with maths teaching, teaching metacognition, teaching problem solving
- Practitioner background information (including experience, maths CPD undertaken, own maths skills and confidence with maths)

As the Stepping Stones theory of change expects practitioner knowledge, confidence and teaching practice to change as a result of taking part in the training, the baseline and endline surveys of Year 2 teachers will include measures of knowledge, confidence/self-efficacy and usual practice for teaching metacognitive practices and teaching problem solving in Year 2. These scales will be developed as part of the project. Confidence and knowledge scales will use multiple items related to the programme content and background and a 5-point Likert scale to indicate level of agreement with the items. Usual practice items will use a five-point frequency scale for how often a teacher engages with different practices related to teaching metacognition. Knowledge related items will ask teachers to indicate their level of knowledge about different aspects of metacognition. While a search of the literature reveals no existing scales which measure knowledge, confidence and practice for teaching metacognition, Mulholland et al. (2025) describe the use of survey items for measuring metacognition in maths in their previous work and similar types of scales in different teaching areas have been used in other evaluation projects (e.g. Fitzfussell et al, 2025). Scale items will be derived from the previous items used in Mulholland et al. (2025) and also derived theoretically from the metacognition literature and based on the theory of change for the Stepping Stones programme. Piloting of the proposed items will be conducted with a small number of Key Stage 1 teachers (including teachers from the pre-pilot and other teachers recruited from researchers' networks) to establish acceptability, face validity and to gain feedback on the items to be included (up to 10). Principal components analysis will be undertaken to understand the structure of the measures and explore how well each of the items work before deciding on the final items to be used in the analysis of the scales.

The surveys will also contain quantitative and more open-ended qualitative questions regarding practitioners' usual practice regarding maths teaching, teaching of problem solving and CPD to investigate the context of the delivery of the programme. This will allow the study to look at how Stepping Stones is different to usual practice, any activity that Stepping Stones will be replacing and how the programme fits with use of other maths programmes or schemes. The baseline teacher survey will also ask about the number of years the teacher has been teaching (categorised as new teachers (0-3 years), mid experience teachers (4-8 years) and very experienced teachers (more than 9 years), teacher's role in the school (teacher only, middle leadership, senior leadership) and the teacher's rating of their own maths skills and confidence (using Likert scale questions). This information will allow us to describe sample of teachers involved and consider whether these factors affect implementation (in terms of number of number of sessions delivered and fidelity of sessions delivered), views of the programme and confidence change (do existing confidence levels affect reported confidence gains?). Endline surveys will also capture information about programme delivery (including number of Stepping

Stones sessions completed, fidelity to the programme materials and core components, adaptations made) and open-ended questions about perceived programme impact (for the teacher, pupils in general, disadvantaged pupils and pupils with SEND). The endline surveys will explore when in the school timetable Stepping Stones lessons have been implemented, what they have replaced, and how they fitted with other maths teaching.

Cost evaluation questions will also be added to the endline surveys to capture the time teachers have spent engaging with Stepping Stones outside of the training sessions (planning and delivery time) as well as any other resource costs they had in delivering the programme.

Baseline surveys will be completed at the start of the first Stepping Stones training sessions to maximise response rates. At the end of the programme, personalised invitations to complete the survey (using the teachers' names and email addresses), will be sent by email to all teachers who completed the baseline survey. The links provided will be specific to each teacher allowing monitoring of completion and reminders to be sent directly to teachers that haven't responded.

The data collected in the surveys relies on teacher self-reporting and recollection of practice. There are some limitations with this method of data collection. Self-report data can be affected by social desirability bias, with participants responding more positively than is accurate. Asking participants to describe practice significantly after the event can also lead to inaccuracies in recall. In terms of reporting confidence and practice, it is also possible that practitioners report more positive practice and confidence when they have less knowledge of the phenomenon, but when they understand more, they report lower confidence and knowledge as they are now more aware of what they don't know or are not doing. While these limitations may limit the conclusions that can be drawn by self-report data, it is still a valuable source of data of teachers practice, understanding and confidence which is it not possible to capture in other ways or where the burden or cost of collecting data in other ways is too much. In this study, triangulation of survey data with other data sources will be undertaken when possible, including triangulation of endline survey response with interview data collected closer to implementation and triangulation of observations of lessons and practice with reported practice where possible for case study setting and initial lesson observations for all schools.

Post-training teacher surveys

All attending teachers will be asked to complete a short online survey to capture knowledge gains regarding metacognition and maths teaching, practitioner confidence to deliver the programme, acceptability of the training, and any feedback practitioners have about the training session. It will also ask teachers to describe any resource needed by the school for them to attend the training (travel cost/distance, provision of supply cover etc.). After the second and third training sessions the survey will ask practitioners to identify the number of Stepping Stones lessons they have delivered up to that point.

Most survey items will collect quantitative using Likert scale responses but there will be the option to elaborate or explain answers by adding a qualitative comment. An open-ended question will also allow practitioners to provide any more general feedback they have about the session. The survey items will be developed in conjunction with the delivery team to make sure we capture necessary information and feedback so that they can understand and refine the training moving forwards. The evaluation and delivery team will agree items for data sharing before the survey is used.

Senior leader baseline and post-intervention surveys

A senior leader in each participating school will be asked to complete a short online survey, at baseline and post-intervention. This survey is designed to collect school contextual information including motivations for signing up to the programme, school organisation and staffing, information about usual practice, use of other maths programmes, resources, schemes of work, resourcing and costs for the programme. At endline the survey will also collect information about the value of the programme for the school, any plans for further embedding of the programme into practice and the acceptability of the cost for the school. This survey will be emailed to senior leaders in advance of the first teacher training session and after delivery of the programme.

Fidelity measures

Lesson observations

Observations of programme delivery across all 15 participating schools will be collected by the programme mentors (part of the delivery team) when they are visiting the school during weeks two or three of delivery. The evaluation team and the delivery team will work together to develop a rubric to be used for these observations. This will be based on the programme implementation checklist developed previously by the delivery team and described in Mulholland et al (2025) to assess the degree to which key programme features are implemented (using a 3-point scale of 1=not implemented, 2=partially implemented, 3=fully implemented).

The items in this scale were:

- 15-20 minutes session delivered, outside of daily maths lesson
- Programme slides used
- Problem solving objective shared with the class
- Mixed-attaining trios/pairs are in place for discussion
- Expectations for discussion are clearly communicated to children
- Strategies for problem solving – in addition to mathematical content alone – are a key focus of the session
- Teacher overview used to inform questioning and discussion within the session
- "River" used
- Alternating worked example used

The evaluation team would also like to include in the rubric additional items reflecting teacher practice and child engagement with the lesson to include:

- Teacher models use of metacognitive strategies during session
- Children actively discuss the problem during the lesson

The rubric would provide examples of the different levels of engagement with each aspect. This would support increased reliability in the application of the measure and also facilitate feedback discussions during the lesson observation by the delivery team early in delivery.

Observations in case study schools will involve the use of this rubric at three time points during delivery conducted by the evaluation team.

Training for the researchers conducting the observations will take place through a joint visit between the researcher and the delivery team mentors during their first visit to the case study

schools. The delivery team mentors are expected to be more experienced in observing for the required elements of the programme due to their knowledge and experience of using it and providing Stepping Stones training. At the end of the visit, the researcher and delivery team mentor from both teams will compare responses and discuss any discrepancies to support the evaluation team researcher's understanding of quality delivery and improve the quality of the observations conducted without the mentors later in the programme.

While observation of the sessions by delivery team members may affect the practice of teachers, likely increasing adherence to the Stepping Stones elements, it is unlikely the observation by the researchers without the delivery team would be free from this bias. The value of being able to collect this data across all settings using the delivery team mentor visits which would not be able to be resourced by the evaluation team, the minimisation of the burden for settings of having one observation visit attended by both evaluation and delivery team, and the opportunity for training and improving reliability and validity of the measure means it is worth using this data.

Quality of teacher-generated materials during training sessions

During training sessions two and three, teachers work collaboratively in pairs to plan Stepping Stones lessons using partially completed templates during training session 2 and blank templates during training session 3. Analysis of these materials generated during the sessions allows insight into how well teachers understand the programme and the capacity of teachers to independently generate these materials beyond the programme. The delivery team review the teacher generated materials using a checklist described in Mulholland et al (2025). During the review of materials, a member of the delivery team rates the teacher generated materials for level of inclusion (detailed information included, information included or whether the information is missing.

The items on this checklist are:

- Learning intentions, including both National Curriculum and problem-solving objectives
- Teacher overview, including script
- Teacher overview includes the use of 'Think Aloud'
- Alternating worked example identified as an optional extension
- Teacher overview details key question prompts
- Worked example uses complete/mistakes/fading as appropriate
- Visual representations are used to support the understanding of key mathematical concepts
- Worked example included in sessions slides includes reference to/exemplification of the problem-solving objectives taught during the session.

For this pilot study the checklist will be developed further into a more detailed rubric with input from the delivery team and evaluation team.

The evaluation team will explore the data by the rubric to look at the different elements of the programme and how well each of these elements has been implemented into the materials. This will be done by looking at the percentage of teacher generated materials which include each element of the checklist the extent and quality to which this is completed.

Where specific elements are not being regularly included in the materials, we will ask teachers during the end of project interviews about how they feel these elements fit into the programme, and their views of the possible reasons for these not being included as part of the teacher

generated materials. These will not be linked to the specific materials generated by that teacher but will explore teachers views of the topic more generally and whether all elements still need to be included for later sessions. Teacher views on the content of the teacher generated materials will be analysed as part of the interview data analysis including the teacher generated materials as part of the coding framework for the framework analysis.

Number of Stepping Stones sessions implemented

To look at the dosage of the programme, post-training surveys after Training session 2 and 3 and Endline teacher surveys will ask teachers to identify to identify how many Stepping Stones lessons they have delivered at each point. Teachers will be asked to explain any reasons for missing sessions.

Longitudinal Case studies

We will purposively sample five schools for in-depth case studies to capture the Stepping Stones delivery experience across different contexts and across the time period of delivery. Settings will be selected before the programme starts, making sure to capture a range of different setting contexts based on their characteristics. Due to the small number of schools in the sample it may not be possible to select settings representing all contexts but we will aim to select at least one large school, one small school, one school with a mixed age class, three schools in deprived areas, one MAT school and a range of Ofsted ratings across the sample. The aim is to see how the programme can be implemented across a heterogeneous sample.

Each school will be visited three times to observe Stepping Stones lessons at different stages of the 'fading' professional development. During the visits researchers will also conduct short interviews with school staff and hold focus groups with Year 2 pupils see Table **5**.

Table 5. Case study schedule

	Visit 1 (Jan/Feb)	Visit 2 (Mar/Apr)	Visit 3 (Apr/May)
Stepping Stones mentor debrief observation	X		
Stepping Stones lesson observation	X	X	X
Class teacher interview	X	X	X
Senior leader interview	X		X
Teaching assistant interview (when applicable)	X		X
Pupil focus group	X		X

Interviews will take place face-to-face using a semi-structured interview schedule and are expected to take between 15 and 20 minutes. Subsequent interviews will draw on ideas generated from analysis of early case study visits and interviews (see later analysis section for description of analysis). All interviews will be audio recorded and transcribed with the transcription used for analysis. Pupil focus groups will take place in groups of 4-6 pupils. Teachers will be asked to recruit pupils with a range of characteristics ensure that disadvantaged pupils, SEND pupils, pupils with strong and weak maths skills are represented in the sample. Ideally, the same pupils will be involved in both focus groups at the different time points.

Pupil Focus Groups

Focus groups will be specifically tailored in their approach to be developmentally appropriate to the 6/7-year-old children. This will include using activities to build trust, establish rapport and create a safe space for conversation. Restricting the group size to up to 6 children should facilitate greater engagement from participant in the group. An icebreaker activity will be used at the beginning of the session (e.g. create a name badge and tell the group something about you) to set the scene for conversation and allow the group to share something safe about themselves helping to establish initially communication. Ground rules would use familiar language to the children (e.g. be good listeners, be kind to each other, be honest) to support understanding of their role in the focus group. Questions and activities would start simple with yes/no answers encouraging participation before moving on to more complex questions. Where possible, topics will be explored through an activity – e.g. add a sticker to how they feel about a statement which can facilitate the conversation. Question wording will be piloted with a small group of Year 2 children to ensure that they are appropriate for the children's level of understanding. The questions and approach to the focus group will be shared with the teacher before the session for feedback and tailoring to the group of children. It is anticipated that pupil focus groups will cover the following topics:

- Feelings towards maths and problem solving
    - Enjoyment
    - Competence
- Enjoyment of Stepping Stones lessons
    - Likes/dislikes
    - What they have learned
- Comparison of Stepping Stones lessons with other maths lessons
- How they approach problem solving in maths
- What they know about meta-cognition and how they use it

Observations of training

Semi-structured observations of all teacher CPD conducted across the programme will be undertaken. This will allow the evaluation team to understand what the training entails and the expectations of school delivery as well as the extent to which the CPD sessions are consistent with the training plan and prespecified materials considering whether the training is adequately defined for scale up beyond the delivery team. The observations will also capture levels of engagement with the training, any questions that teachers ask during the training, reflections on delivery including successes and barriers to delivery. Observations of mentoring visits will also look at consistency of support offered across the visits.

Interviews

All interviews outside of case study visits will be conducted remotely using Microsoft Teams. Recordings will be made with permission of the participants and transcripts produced for analysis.

Teacher interviews (post-intervention)

To explore the perceptions of a wider range of teachers, remote interviews (using Microsoft Teams) will be conducted with a teacher in each of the non-case study participating schools at the end of the intervention delivery period (n=10). Semi structured interview schedules will be used to explore Stepping Stones implementation, any barriers or facilitators for implementing the programme and perceived impact for staff and children. Engagement, barriers, facilitators

and impact for disadvantaged pupils will be specifically included in the interviews. They will also explore topics arising during the case studies to investigate the generalisability of these themes across the wider sample. Resourcing and time spent delivering (and planning for delivering) the Stepping Stones will also be explored.

Teacher interviews (follow-up)

Follow-up semi-structured interviews will be undertaken remotely (using Microsoft Teams) with a teacher in all participating schools in September/October 2027 (n=15). These interviews will explore the longer-term use of teaching metacognitive strategies in the school beyond the programme. This will include whether the school plans to implement the Stepping Stones lessons again with Year 2 pupils, whether teachers are using any of the Stepping Stones strategies in their normal maths teaching or teaching in other lessons and whether Stepping Stones practice has spread beyond the teachers who were involved in the programme.

Delivery team interviews

Two group delivery team semi-structured interviews will be conducted: the first during programme delivery (March 2027 after delivery of two training sessions) and the second after completion of the project (June/July 2027). The interview topics will include:

- reflections on the delivery of the professional development through the programme
- programme implementation in schools
- any adaptations made to the programme through delivery
- scaling the programme up for delivery beyond the pilot.

Piloting of metacognitive thinking assessments

A specific request of the EEF for this project is the identification of appropriate outcomes for measuring metacognitive thinking that could be used in a subsequent efficacy trial. Review of the current literature identified few measures of metacognition that have been used with this age group particularly and no established tools for measuring metacognition in maths and problem solving.

Assessments of metacognition are categorised either as 'online' – metacognition measured while engaged in a learning task – or 'offline' – retrospective reports about previous use of metacognitive strategies. Online assessments tend to involve audio or visual recording of the performance of a participant in a task and coding of their behaviours and language use. While online assessments are more burdensome and costly to implement, online assessments with school students have better predicted course performance in a previous study albeit with older students while offline measures did not (Treglia, 2018). Developmentally, at age 6 and 7 pupils are not expected to be able to reflect and describe their metacognitive processes offline, although are able to develop and use metacognitive strategies in their work which would be captured by an online assessment. In this age group, some previous studies have used offline teacher rating scales for offline measurement of metacognition and we would also trial this less burdensome method.

Through the study we will develop and test two metacognition assessment measures designed to capture Year 2 pupils' metacognitive behaviours during mathematical problem-solving: an online 'Think Aloud' assessment and an offline teacher rating assessment. We will be investigating the feasibility and acceptability of the data collection for these approaches, the suitability of the measures for capturing young pupils' metacognitive thinking, and their potential for use in a future trial. We will also look at how the two measures correlate with each other and with children's maths ability as rated by the teacher.

‘Think Aloud’ Assessment

The online pupil metacognition measure is expected to be a think-aloud problem-solving task, in which pupils verbalise their reasoning while solving carefully designed mathematical problems. The pupil verbalisations are seen as providing insight into the thinking processes of the child about the maths problem. Audio recordings of pupils' verbalisations along any annotations or shown working on the maths problem sheet, will be coded using an adaptation of Desoete’s (2001, 2002, 2008) validated metacognition coding framework, which has previously been used effectively with primary-aged pupils.

The coding scheme assesses four components of metacognitive regulation:

- Prediction (11 items): e.g., identifying what is known, rereading to clarify, estimating outcomes
- Planning (6 items): e.g., selecting relevant steps, choosing operations, devising an action plan
- Monitoring (17 items): e.g., checking steps, adhering to plan, verifying calculations
- Evaluation (6 items): e.g., summarising, reflecting on success, generalising to future problems

Each behaviour receives a score of 0 (absent), 0.5 (initiated) or 1 (present). This creates a structured, fine-grained measure of pupils’ metacognitive engagement.

Capturing students' verbalised problem-solving processes allows direct online capture of real-time metacognitive processes, which cannot be reliably inferred through written work alone or through reflection after completing the problem (especially not in this age group). The ‘Think Aloud’ method also aligns with the Stepping Stones programme strategies and with the expected change to pupil ability to verbalise, plan and regulate their learning. There are some considerations that will need to be addressed through the project regarding the use of this type of assessment. The action of asking pupils to think aloud is a meta cognitive strategy in itself and may introduce thinking and use of strategies that would not have be used otherwise. Also, the suitability and acceptability of this type of assessment for all children needs to be considered through the project especially with consideration of those with low maths skills or who are less confident verbally.

<u>Pre-piloting assessment development phase</u>

A pre-piloting phase will test and refine the suitability of the problem-solving question for the task (difficulty, accessibility, alignment with Year 2 curriculum), the prompting and instruction that will be given to the child during the task, the clarity of audio-recordings for reliable transcription and the views of teachers on delivering this assessment. It will also look at how many questions should be used and whether the exercise works better as an individual task or a small group task (pairs or threes).

Following consultation with maths, metacognition and assessment development experts, a bank of questions will be developed for trialling in five schools. This will be undertaken with a Year 2 class teacher or teaching assistant and four to six pupils per school in five schools not involved in the main Stepping Stones pilot. We are designing the assessment to be undertaken by teachers or teaching assistants and in the pre-pilot we will directly work with these school staff to develop guidance and instruction for delivery. Pre-pilot development with schools will be an iterative process with experiences in one school informing the trialling experience in the

next school. Exploration of the feasibility of the delivery for the main pilot will also be explored with teachers including required resourcing of staff, use of quiet space for recordings and practicality of the audio recorder use.

Main pilot delivery

The main pilot will test the developed assessment approach with a wider sample of pupils and using a fully remote model. Five schools will be selected to trial the metacognitive assessment 'Think Aloud' measures developed as part of the pre-pilot stage with assessment of 10 pupils per school. Teachers or teaching assistants will be asked to conduct the assessment task with the pupils in their class. An audio recorder and materials will be sent to the school in the post in advance of the testing period or provided at the final training session. Training for school staff in assessment delivery will be provided via a webinar before the testing period and full instructions given in the pack sent to schools. The teacher or TA delivering the assessment will be asked to provide the maths problem to the pupil (or small of pupils), ask the pupil to 'Think Aloud' while solving the problem, provide prompts to encourage the verbalisation of the pupils and to ask specific debrief questions at the end. The full process will be audio recorded by teachers and the recordings and prompts sent back to the evaluation team for coding. Background information about pupils being assessed (maths ability, pupil premium eligibility and gender) will be collected from the school for each pupil. We will also ask teachers to complete a short feedback form about the assessment process to include:

- Time taken to complete the assessment
- Suitability of the questions for pupils of different abilities
- Suitability of the questions for getting students to 'Think Aloud'
- Whether the task accurately reflected the metacognitive ability of the pupils
- Any challenges with conducting the assessment or anything that was unclear
- Reflections on the assessment process.

Transcripts will be made of audio recording data and coding undertaken using anonymised transcripts. Coding of the assessment data will be undertaken by two coders at the end of the project, supervised by Linda Wang. A training period will be used where all coders code the same audio recordings with discussion of any discrepancies following the coding of each recording and clarification of the codebook made to aid future decisions. This will be repeated with different recordings until coders reach at least 90% reliability with each other applying the code book. After the training period 10% of recordings will be double coded. We will also explore the use of AI comparative judgement for these assessments to assess the suitability for more large-scale data processing. We expect to do this through use of No More Marking's AI Comparative Judgement system. This system uses an established platform where human judges compare two side-by-side pieces of student work and choose which is better based on predefined criteria. The platform now integrates AI judges (using OpenAI) to execute these pairings. Instead of asking an AI to blindly assign an absolute grade, the system uses a hybrid model: the AI handles a given percentage of the repetitive comparative decisions, while human judges can also complete judgements to validate the AI. This approach not only reduces traditional judging times, but it also eliminates common human biases. Exploration would be conducted of the feasibility and capability of this system to judge use of metacognitive strategies in pupil transcripts during the 'Think Aloud' task.

Piloting of Teacher rating scale of child's metacognitive ability

During the pre-piloting period we will also develop a teacher rating scale of children's metacognitive thinking. While the 'Think Aloud' task described above is a measure of 'online' metacognition as it is measuring the strategies and behaviours actually used during a task, a rating scale can give an off-line measurement of metacognition which asks for reflection of overall ability and behaviour rather than task specific behaviour. A previously-developed teacher rating scale of metacognition has correlated with maths ability and other metacognition measures with older children in primary school (Desoete, 2008). Use of a teacher rating scale rather than an active assessment which required an involved process from teachers plus later coding, would minimise cost and data collection burden in a wider trial if it could be shown to be a valid method of data collection. We will draw on and adapt the previously used teacher report scales considering the suitability for the 6-7year age group and the UK context.

Teachers in the five assessment pilot schools will be asked to complete the teacher rating metacognition scale for the children who do the 'Think Aloud' task. It will be completed in advance of assessing children using the 'Think Aloud' method so that their judgements are not influenced by their delivery of this assessment. Teachers will also be asked to reflect on the acceptability and relevance of the questions in this scale, the time taken to complete it and relevance to the children in their classroom after they have completed it and will provide details of this in the proforma. The previously described post-interview interviews will also explore teacher experience with the rating scale and the 'Think Aloud' task.

We will look at the relationship between the teacher rating scales and the 'Think Aloud' scoring to investigate whether they are both tapping into the same constructs.

Cost evaluation data collection

Due to the pilot nature of the project a full cost evaluation will not be undertaken, however the project will gather data on the costs of a potential future scaled up programme which could be delivered as a part of a trial and collect data on the acceptability of future programme costs for schools. Data will also be collected from schools on the time requirements for staff to undertake the programme and any additional costs schools incur for taking part in the programme beyond the provision of the PD and materials.

A Cost and Scale up workshop between the delivery team and evaluation team will explore any adaptations that would need to be made for scaling up the programme for delivery beyond the pilot and explore the costs that would be required for providing the Stepping Stones programme to schools in this model.

Senior leaders will be asked to rate the acceptability of the scaled-up programme costs (explored in the cost and scale up workshop) in the endline survey. This would also be discussed with senior leaders in case study settings.

Time and any additional resources/costs required to implement the Stepping Stones programme during the pilot delivery will be explored in the teacher interviews at endline and with the case study settings. Post-training surveys will ask teachers about whether additional supply cover for required for them to attend the training sessions.

Summary of data collection methods

Table 6 below provides a summary of the data collection methods that would be used during the study and how this data relates to the research questions. Full details of the data analysis methods are described later.

Table 6. Data collection methods

Research Methods	Data Collection methods	Evaluation dimension	RQs addressed (Success criteria)	Sample size and sampling criteria	Data analysis methods
Administrative data	Training attendance and delivery records	Feasibility and Acceptability	6 (SI6)	Records for all who attend training	Descriptive statistics for number and percentage of teachers at each training session.
	Fidelity records from coaching visits to schools	Feasibility and Acceptability	7 (SI5)	15 schools (collected and provided by DT)	Descriptive statistics from fidelity rubric. Framework analysis applied to qualitative data.
	Data available about schools from online sources e.g. Get information about schools and school websites.	Feasibility and Acceptability	6, 7, 8, 9, 10 (SI5, SI6, SI8)	As available – data sought for all settings	Descriptive statistics
	Programme materials	Readiness for trial	11, 12 (SI9, SI10)	As available from developer	Document analysis
	Delivery team analysis of created lessons plans by teachers for lessons	Feasibility and Acceptability, Readiness for trial	1, 2, 6, 7, 11 (SI5, SI9)	As available from developer	Descriptive statistics

Senior Leadership survey (baseline/ endline)	Online questionnaire to capture setting practice and context, reasons for signing up, views of programme, impact and acceptability of cost.	Feasibility and Acceptability, Evidence of promise, Readiness for trial	1, 2, 3, 4, 7, 8, 9, 10, 13 (SI3, SI8, SI9)	15 school senior leaders/head teachers	Descriptive statistics, Framework analysis for qualitative data
Teacher survey (baseline/ endline)	Online questionnaire to capture confidence, knowledge and practice in teaching meta-cognitive strategies for problem solving and intervention techniques, as well as other similar training. At endline, survey will also capture information about Stepping Stones delivery and views of the programme. Inclusion of quantitative measures and open-ended questions.	Evidence of promise, Feasibility and Acceptability	1, 2, 3, 5, 7, 8, 10, 11 (SI1, SI2, SI3, SI4, SI5)	All Year 2 teachers 15 schools (expected 15-30)	Descriptive statistics, comparison of scale scores from baseline to endline, framework analysis for qualitative data
Post-training Surveys	Short online questionnaire after each training session to capture knowledge	Evidence of promise, Feasibility and acceptability, Readiness for trial	1, 6, 11 (SI1, SI5, SI9)	All Year 2 teachers who attend training (expected 15-30)	Descriptive statistics, framework analysis for qualitative data

	gains, confidence in programme delivery, acceptability of the training, feedback on training, and resource costs to attend				
Observations of training	Semi-structured observation of training for teachers to capture engagement with training and consistency of delivery	Feasibility and acceptability, readiness for trial	6, 9, 11, 12 (SI6, SI8, SI9, SI10)	3 in person training days plus up to 3 online support sessions	Framework analysis applied to observation field notes.
Observations of practice (case study settings)	Semi-structured observations of Stepping Stones lesson delivery at three time points during the project and developer team in person support visit.	Evidence of promise, Feasibility and acceptability, readiness for trial	2, 3, 7, 8, 9, 10, 11 (SI2, SI5, SI7, SI8)	5 case study settings selected to represent a variety of characteristics with observations at 3 time points	Framework analysis applied to observation field notes.
Interviews/ focus groups	Semi-structured phone/remote interviews (end of intervention)	Evidence of promise, Feasibility and acceptability, readiness for trial	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 (SI1, SI2, SI3, SI4, SI5, SI7, SI8, SI9)	10 Year 2 teachers one from each non-case-study setting	Framework method using broadly inductive approach.
	Semi-structured teacher interviews (three time points) in person during	Evidence of promise, Feasibility and acceptability, readiness for trial	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 (SI1, SI2, SI3, SI4, SI5, SI7, SI8, SI9)	5 Year 2 teachers across 5 case study settings t at three time points during delivery.	

	case study visits				
	Semi-structured senior leader interviews (two time points) in person during case study visits	Evidence of promise, Feasibility and acceptability, readiness for trial	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,13 (SI2, SI3, SI8, SI9)	5 senior leaders across 5 case study settings	
	Semi-structured teaching assistant (interviews (two time points)	Evidence of promise, Feasibility and acceptability,	2, 3, 4, 5, 7, 8, 9, 10, (SI2, SI5, SI7, SI8)	Up to 5 teaching assistants across 5 case study settings	
	Semi-structure phone/remote interviews (term after intervention)	Evidence of promise	2, 5 (SI2, SI3)	15 Year 2 teachers	
	Pupil focus groups	Evidence of promise, feasibility and acceptability	3, 4, 7, 8, 9, 10 (SI7, SI8)	4-6 Year 2 children in 5 case study settings at two time points during delivery.	
	Semi-structured group interviews	Feasibility and acceptability, readiness for trial	6, 7, 8, 9, 10, 11, 12, 13 (SI5, SI6, SI8, SI9, SI10)	2 with up to 4 members of developer team	
Piloting of metacognitive thinking and problem-solving measures	'Think Aloud' problem-solving task and semi-structured debrief discussion	Readiness for trial	14 (SI11)	50 Year 2 children, sample of 10 children in each of 5 assessment schools (not case studies)	Coding according to quantitative coding framework and descriptive statistics. Framework analysis applied to discussion notes.

	Teacher rating scale of metacognitive skill and semi-structured debrief discussion	Readiness for trial	14 (SI11)	Use in 5 assessment schools (not case studies) with recruited Year 2 children	Descriptive statistics for measure. Correlations with 'Think Aloud' task for those pupils. Framework analysis applied to
Cost/scale-up workshop	Semi-structured workshop to establish scaled up delivery models and potential costs	Cost	12, 13 (SI9, SI10)	1 with developer team	Description of scaled up delivery model and table presenting costs split by ingredients

Data analysis

Quantitative survey data

Descriptive statistics will be presented for the quantitative survey questions. Data will also be explored by type of school (MAT school vs local authority school), and by larger class cohort (20+) versus smaller class cohort (>20).

Principal components analysis will be used to explore and analyse the factor structure of the developed self-report scales of practitioner confidence, knowledge and teaching practice measures in the teacher usual practice surveys and to identify any items that are not working as intended. The factor structure model with the best fit both logically and statistically will be used to create scale scores. Cronbach's alpha for scale reliability will be presented for each developed scale.

For each of the developed scales, descriptive statistics will be presented. The total score for each scale will used in a paired samples t-test to investigate the difference in scores before the Stepping Stones Programme and after the programme. Effect sizes (Cohen's D) will also be presented.

Qualitative data from interviews, focus groups, observations and teacher discussions

All data collected in interviews and focus groups will be transcribed from the recordings and analysed using the Framework Method (Gale et al, 2013, Klingberg et al, 2024) through a multi-step process. using a broadly deductive approach with some elements of inductive coding. The framework method provides a systematic, transparent and flexible process for conducting

thematic analysis. It involves organising data in a structured matrix to facilitate comparison within and across cases.

We will follow the guidance set out in Gale et al (2013) to conduct this analysis. Transcripts produced by an external transcription company will be checked by listening to the audio and comparing it with the transcript, to ensure transcript accuracy and to familiarise the researcher with the data. For observation field notes and teacher discussion notes familiarisation will be done through rereading notes. Following familiarisation, the researcher will code each meaningful line of the transcript applying a label or code describing what has been identified as in important in that passage. A broadly deductive approach will be taken at this stage using pre-defined codes related to the study research questions and theory of change for the programme. At this stage, open coding will be used for elements that do not align with the pre-defined codes. Two researchers will independently code the first few transcripts and will meet to compare the labels and agree on the codes to apply to the other transcripts. The two researchers will agree a working analytical framework grouping codes together where necessary and incorporating any unexpected codes into the framework. An 'other' code will be used when data does not fit into the existing framework. This framework will then be applied to the coding of subsequent transcripts which will be undertaken by one researcher. Regular meetings with the second researcher will discuss the 'other' codes, allow reflections and discussion of the process and may lead to updates to the analytical framework and the development of interpretative notes. Once the framework has been applied to all transcripts, a spreadsheet will be created cases as rows and codes as columns. Data in each coding category will then be summarised for each transcript filling in the matrix. The framework will then facilitate the interpretation of the data which the researcher will review and make connections within and between participants and categories guided by the research questions and themes generated inductively from the data. Ideas generated from this interpretation will be written as analytical memos and discussed with the other researcher and team as a whole to refine interpretation. Findings will then be described through the research questions and linked to the programme theory of change. Throughout the process the researchers will use a reflective diary to track analytical choices, capture interpretive thoughts and to consider the reflexivity of the researchers.

Data will be initially analysed after each data collection period and so that it can inform interview/focus group topics at later data collections. Insights from the qualitative analysis may also inform end of project survey questions. At each point of data analysis, the analytical framework will be revised as required with new labels generated from later data. Coding and the generation of the framework will be conducted using Computer Assisted Qualitative Data Analysis software (CAQDAS).

Training and Mentoring visit observation data

Observation data from training sessions and mentoring sessions will be analysed deductively within the framework to look at the extent to which training followed the pre-specified structure and content, consistency between mentors in mentoring sessions, fidelity to the programme and engagement of participants in the training/mentoring sessions. Analysis will also look for any participant-described barriers and facilitators to implementation.

Lesson observation rubric fidelity measure

Descriptive statistics for each of the items in the lesson observation rubric measure will be reported to explore the variation in fidelity for each of the core elements of the Stepping Stones programme across all participating schools. This will be presented for all 15 settings with the data from the delivery team mentoring visits, and for the five case study settings involved in the evaluation team observations with fidelity over the three time points presented for these settings. Exploratory analysis will investigate whether any aspects of fidelity vary for schools with mixed classes and by size of Year 2 cohort. Fidelity scores will also be created using these checklists to explore an overall fidelity score for the setting which could be used in a future trial. End of intervention interviews use the data from these observations to explore with teachers their reasons for not implementing certain aspects of the programme.

Quality of teacher-generated materials fidelity measure

Descriptive statistics will be presented for each of the items in the lesson plan checklist to provide an overall description of the quality of materials created by teachers. As this data is created collaboratively it will not be broken down by individual teacher or school. A ranking of the elements with the greatest fidelity to those will the lowest fidelity will be presented. Items with low fidelity will be explored in endline teacher interviews to look at reasons for them not being consistently included.

Piloting of metacognitive thinking assessments

Audio recordings of the 'Think Aloud' assessment will be transcribed and the transcripts, along with the problem worksheet, will be coded using the described coding framework. Descriptive statistics will be presented for each of the coding categories to show the types of metacognitive behaviours that are highlighted by the assessment. Inter-rater coding reliability will be presented. The average time taken to complete the assessments per pupil will also be presented. Level of agreement between AI comparative judgment scores and the researcher coding will also be presented.

For the teacher rating scale, descriptive statistics will be presented for the overall scores and the individual items in the scale. Principal components analysis will be undertaken to look at the factor structure of the measure and the reliability of the items within the measure. Correlations between the 'Think Aloud' assessment scores and the Teacher rating scale will also be presented.

Deductive analysis of the qualitative responses given in the teacher proforma regarding the acceptability, suitability and feasibility of the assessment process will be undertaken.

Specific consideration of disadvantaged pupils accessing and experiencing the programme

Research question 9 specifically addresses how disadvantaged pupils engage with the programme. This will be explored through analysis of pupil focus group data (which will include disadvantaged pupils in the sample) and through teacher interview data with specific questions asked designed to encourage reflection on how disadvantaged pupils access and engage with the programme including any specific barriers and facilitators for engaging disadvantaged pupils. Supporting disadvantaged pupils is also expected to be explored at the teacher PD days and reflections of participants captured during observations of these sessions will also be analysed. The pre-developed code book for analysis of the qualitative data will specifically

include disadvantaged pupils so data will be analysed deductively looking specifically for participants responses regarding this theme.

Cost evaluation

Using the ingredients model from the EEF's cost evaluation guidance (2023), we will present the potential costs for a scaled-up model of the Stepping Stones in terms of the cost per setting for the first year of intervention delivery.

The cost evaluation will also present the time required for school staff to deliver the programme including attending training, planning and lesson delivery during the pilot programme and any additional costs that schools incurred in doing the Stepping Stones programme during the pilot.

Ethics and registration

Ethical approval

Ethics approval for the evaluation has been sought from the School of Education Ethics and Data Protection committee separately for the pre-pilot activities (approval granted 7th November 2025) and for the main pilot evaluation (approval granted 18th December 2025). At the time of approval, not all data collection tools had been drafted and these will be submitted for additional approval before use. Ethical approval for the delivery of the Stepping Stones programme will be sought by Northumbria University.

Agreement to take part

Informed, opt-in consent will be obtained from settings, parents for child participation, and directly from adult participants for each research element.

School senior leadership will be provided with an information sheet and will sign a Memorandum of Understanding (or participation agreement in the case of pre-pilot schools) when agreeing for their setting to take part. Informed consent will be sought directly from adults in the study for each research element through the provision of information sheet and verbal summary when appropriate before undertaking the research. For child participation in case study focus groups and piloting the metacognitive assessments, schools will be asked to inform parents about the study through an information sheet and to collect opt-in consent forms for the child's participation in the research. Child assent will also be sought through verbal briefing and response before the child undertakes any research activities.

Data protection

All information collected as part of this project relating to participants will be treated in the strictest confidence and processed in accordance with General Data Protection Regulations (GDPR). Only data required to run the programme and conduct the evaluation will be collected and all data will be collected, stored and shared securely with access only to those working on the project.

The lawful basis the evaluation team are relying on for collecting and processing data as part of this project is 'public task', which according to General Data Protection Regulations (GDPR) Article 6 (1) (e) means 'the processing is necessary for an activity being carried out as part of the

University's public task, which is defined as teaching, learning and research'. This project is carrying out research and does not intend to process special category personal data.

Durham University and Northumbria University will act as data controllers or joint data controllers for the project. Data sharing agreements will be put in place between Durham University and Northumbria University to manage the sharing of data for the project and between Durham University and the participating schools for sharing data about participating pupils for the project.

A Data Protection Impact Assessment (DPIA) was conducted for this project by the principal investigators, describing why the collection of personal data is necessary, the data that will be collected and the ways that risks will be minimised and this has been reviewed and approved by Durham University Information Governance. The data retention period for storing and processing personal data is five years beyond the end of the project (July 2033) to allow for the verification of data for further publishing of papers beyond the report. At this point, personal data will be deleted or anonymised by the Evaluation Team. Anonymous data may be kept indefinitely and potentially shared with other research teams for future research.

Data will be stored and processed in line with Durham University's Data Management Policy. A data management plan has been written by the PI and will be implemented by the evaluation team to manage the handling, storing and sharing of project data.

Participants (setting staff and parents/carers) will be provided with information sheets and privacy notices before agreeing to take part in the project which set out what, and how, data will be collected and processed as part of the project, how it would be stored, who it would be shared with along with information about participants ability to withdraw from data processing as part of the project and the retention period for project data.

Personnel

Delivery team

Dr Kirstin Mulholland	**Associate Professor, Communities and Education**	**Northumbria University**

Kirstin co-developed the Stepping Stones programme, and will lead the delivery team. Her roles include co-designing and delivering programme training, and providing in-school support for participating teachers. Kirstin is also responsible for the recruitment of schools for this project, and will be the main point of communication for the larger delivery team.

William Gray	**Deputy Headteacher**	**Westerhope Primary School**

William co-developed the Stepping Stones programme, and is the deputy-lead for the delivery team. Williiam is responsible for co-designing and delivering programme training, and providing in-school support for participating teachers. He has also supported the recruitment of schools for this project.

Dr Chris Counihan	**Assistant Professor, Communities and Education**	**Northumbria University**

Chris provides ongoing support for the design and delivery of the Stepping Stones programme, with a particular emphasis on evaluation of the programme training. He supports programme co-ordination, logistics, and recruitment.

Emma Barker	**Deputy Headteacher**	**Amberley Primary School**

Emma provides support to the delivery team, including designing programme materials and providing in-school support for participating teachers. Emma has also supported the recruitment of schools for this project.

Dr David Nichol ***Assistant Professor, Communities and Education*** ***Northumbria University***

David provides ongoing support for the design and delivery of the Stepping Stones programme, with a particular emphasis on evaluation of the programme training. He supports programme co-ordination, logistics, and recruitment.

Evaluation team

Vic Menzies **Assistant professor, Education** **Durham University**

Vic will be co-Principal Investigator, leading the evaluation design, protocol development, project delivery, reporting, and management. She will also conduct data collection and analysis.

Dr Linda (Yuqian) Wang **Associate Professor, Education** **Durham University**

Linda will be co-Principal Investigator and Mathematics Education expert. She will lead the development of measures for metacognition in Mathematics problem-solving and contribute to evaluation design, quality assurance, data collection and reporting.

Dr Xiaofei Qi **Associate Professor, Education** **Durham University**

Xiaofei will advise on IPE and assessment. She will support research design, ensure quality assurance in data collection, and contribute to the development of research tools and measures.

Dr Rachel Oughton **Associate Professor, Mathematical Sciences** **Durham University**

Rachel will be the project statistician, advising on design and measure development, and conducting statistical analysis of quantitative data.

Risks

Table 7 below highlights key risks to the project delivery and how we aim to manage and mitigate these risks. Risk management will be integrated into monthly project team meetings to highlight any upcoming areas that may cause issues, raise and manage them early and ensure successful delivery of the evaluation.

Table 7. Project risk management

Risk	Proposed mitigation
Poor response rates to survey data collection leading to poor quality or insufficient data for answering research questions across the project.	Survey data collection done as part of in person events to maximise response. Paper copies sent directly to schools where no online response received. Surveys kept short. Accessible information provided through short video on value of evaluation activities and expectations of participants at beginning of project. Positive relationships with settings established through regular contact.
Selected case study or assessment settings unwilling to facilitate visits or allow data collection.	Selection of five case study settings allows for potential withdrawal of one setting and still maintain a variety of cases. Selection of case study and assessment settings at the

	beginning of project means expectations are clear for schools from the start. Incentive payments recognise additional workload for schools involved for these aspects of the work. Early phone conversations with case study schools and assessment schools to explain what this will mean and allow answering of questions. Alternative schools could be selected if chosen schools are unwilling from the start.
Audio recordings of meta-cognitive tasks not sufficiently clear for transcription and analysis or audio recording process not acceptable to schools or children.	Pre-piloting of audio recordings processes and coding during set up period will allow early identification of these issues. This would allow sufficient time for explorations of other recording approaches e.g. use of mics or video recordings. Piloting would also flag unacceptability and allow trial of other approaches e.g. group tasks or a cognitive interview or observation approaches.
Insufficient school recruitment or high attrition rates.	Establish clear communication channels with school leaders, provide ongoing support and resources to participating schools, over-recruit by 20%, and conduct regular check-ins to proactively identify and address concerns.
Inconsistent or poor-quality delivery in the classroom.	Classroom observations will be scheduled for 5 case study school visits, followed by feedback sessions. During the pre-pilot period, we will work with the developer to create a structured checklist based on their existing tools.
External validity is threatened by only recruiting a small number of homogenous schools.	We will work to actively ensure recruitment of a diverse sample across different school contexts and document contextual factors that may influence implementation.
Participants' lack of engagement in evaluation may lead to poor quality data collection and risk validity of research findings.	Clear and accessible information sheets and short video make clear evaluation expectations from the start of the programme. Scheduled, regular communication which includes the benefits of trials is expected to improve buy-in. Using incentives for higher intensity research activities allows schools to feel appreciated and leads to greater engagement in evaluation tasks.
Safeguarding and ethics risks	Ensure the team holds enhanced DBS checks and is safeguarding trained in line with Durham University policy. We will implement clear protocols for identifying and reporting concerns from participant teachers or from school visits. Consideration will be made of the burden of the research tasks on participating children and tasks will be kept

	short and management. Initial piloting will allow determination of time spent.
Delivery risks	Ensure clear milestone checkpoints align with the Table 4 timeline. Also, prepare alternative delivery models if face-to-face (f2f) training isn't feasible for some participants.
Budget constraints	Regular meetings will include budget monitoring.

Timeline

The timeline for the project is described in Table 8 below.

Table 8. Project Timeline

Date	Activity	Staff responsible/leading[1]
August to October 2025	Set up meetings	EEF, NU, DU
October to December 2025	Development of recruitment documentation	NU/DU/EEF
December 2025	Ethics application	DU
January to March 2026	Development of data sharing agreement	DU/NU
March/April 2026	Draft Study Plan	DU
February to April 2026	Setting recruitment	NU
February to June 2026	Development and piloting of research tools including surveys and metacognitive measures	DU
September/October 2026	Senior Leader Baseline Survey	DU
October 2026	Selection of case study & child assessment settings	DU
November 2026	1st Stepping Stones training day	NU
November 2026	Year 2 teacher Baseline Survey (as part of training day)	DU
November 2026	Observation of 1st Stepping Stones Training Day	DU
November 2026	Year 2 teacher post-training survey (as part of training day)	DU
January 2027	Observation of Stepping Stones sessions by coaches	NU
January 2027	1st case study visits (observation of practice & coaching, interviews with teacher, SL, TA, pupil focus group)	DU
February 2027	2nd Stepping Stones Training Day	NU
February 2027	Observation of 2nd Stepping Stones Training Day	DU

[1] DU = Durham University, NU = Northumbria University, EEF = Education Endowment Foundation

February 2027	Year 2 teacher post-training survey (as part of training day)	DU
March 2027	1st Developer group interview and cost/scale-up workshop	DU
March to April 2027	2nd case study visits (observation of practice, interview with teacher)	DU
March 2027	3rd Stepping Stones Training Day	NU
March 2027	Observation of 3rd Stepping Stones Training Day	DU
March 2027	Year 2 teacher post-training survey (as part of training day)	DU
April to May 2027	3rd case study visits (observation of practice, interviews with teacher, SL, TA, pupil focus group)	DU
May 2027	Remote practitioner interviews (in non-case study settings)	DU
May to June 2027	Year 2 teacher Endline survey (online)	DU
May to June 2027	Senior Leader Endline survey (online)	DU
May to June 2027	Collection of pupil meta-cognitive data in assessment settings	DU
June/July 2027	2nd Developer group interview and feedback session	DU
July to November 2027	Data coding, cleaning and analysis	DU
September/October 2027	Follow-up remote interviews with Year 2 teachers	DU
October to December 2027	Report writing	DU
December 2027	Submit draft report	DU
January to May 2028	Revise report based on feedback rounds	DU/EEF/NU
May 2028	Submit final report	DU

References

Anh Do, L. & Thomas, A. (2025). In-person collaboration, but not virtual, enhances problem-solving efficiency and knowledge transfer from groups to individuals, *Mind, Brain, and Education,* 19. https://doi.org/10.1111/mbe.70008

Bingham, A. J. (2023). From Data Management to Actionable Findings: A Five-Phase Process of Qualitative Data Analysis. *International Journal of Qualitative Methods*, 22. https://doi.org/10.1177/16094069231183620

Boylan, M., Adams, G., & Birkhead, A. (2023). Landscaping mathematics education policy: Landscaping national mathematics education policy. Sheffield Hallam University.

Branigan, H., & Donaldson, D. (2020). Teachers matter for metacognition: Facilitating metacognition in the primary school through teacher-pupil interactions. *Thinking Skills and Creativity, 38*, 100718.

Chatzipanteli, A., Grammatikopoulos, V., & Gregoriadis, A. (2013). Development and evaluation of metacognition in early childhood education. *Early Child Development and Care*, *184*(8), 1223–1232. https://doi.org/10.1080/03004430.2013.861456

Chen, S., Guo, M., & Dousay, T. A. (2023). Grow to learn: a metacognitive approach to early childhood teachers' science professional development. *Research in Science & Technological Education*, *43*(2), 390–410. https://doi.org/10.1080/02635143.2023.2279076

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities, 34*(5), 435-449. https://doi.org/10.1177/002221940103400505

Desoete, A., Roeyers, H., & De Clercq, A. (2002). Can offline metacognition enhance mathematical problem solving? Journal of Educational Psychology, 95(1), 188-200. https://doi.org/10.1037/0022-0663.95.1.188

Desoete. (2008).Multi-method assessment of metacognitive skills in elementary school children: How you test is what you get. Metacognition and Learning, 3(3), 189-206. https://doi.org/10.1007/s11409-008-9026-0

Desoete, A. (2007) Evaluating and improving the mathematics teaching–learning process through metacognition, *Electronic Journal of Research in Educational Psychology*, ***5****(3),* 705–730.

Desoete, A., Baten, E., Vercaemst, V. De Busschere, A., Baudonck, M., & Vanhaeke, J. (2019). Metacognition and motivation as predictors for mathematics performance of Belgian elementary school children. *ZDM Mathematics Education, 51 (4). https ://doi.org/10.1007/s1185 8-018-01020 -w.*

Department for Education. (2024). *Trends in international mathematics and science study (TIMSS) 2023: National report for england (Vol. 1).* Department for Education.

Department for Education. (2021). *National curriculum in England: Mathematics programmes of Study.* National curriculum in England: mathematics programmes of study - GOV.UK

Donovan, M. S., & Bransford, J. D. (2005). *How students learn–Science in the classroom*. National Academy Press.

Education Endowment Foundation. (2020). *Improving Mathematics in the Early Years and Key Stage 1*. *Guidance Report*. Education Endowment Foundation.

Education Endowment Foundation. (2021). *Metacognition and self-regulated learning: Guidance report*. Education Endowment Foundation

Education Endowment Foundation. (2025). Metacognition and self-regulated learning: Guidance report. London: Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/metacognition/metacognition-and-self-regulated-learning_guidance-report.v.2.4.0.pdf

Fitzfussell, N., Menzies, V., Eerola, P., Daniel, J., Burgoyne, K., Qi, X., & Cramman, H. (2025). *Talk With Tales for Children (TWiTCH) programme*. *Pilot evaluation report*. Education Endowment Foundation.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https:/ /doi.or g/10.10 37/00 03-066X.34.10.906

Gale, N., Heath, G., Cameron, E., Rashid S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology, 13,* 117.

Hitt, L. (2023). A systematic review and meta-analysis of interventions based on metacognition and self-regulation in school-aged mathematics (Doctoral dissertation, Durham University). https://etheses.durham.ac.uk/id/eprint/15035/

Hodgen, J., Foster, C., Marks, R., & Brown, M. (2018). *Evidence for review of mathematics teaching: Improving mathematics in key stages two and three: Evidence review*. Education Endowment Foundation.

Jossberger, H., Brand-Gruwel, S., Boshuizen, H., & Wiel, M. (2010). The challenge of self-directed and self-regulated learning in vocational education: A theoretical analysis and synthesis of requirements. *Journal of Vocational Education & Training*, *62*, 415-440.

Klingberg, S., Stalmeijer, R. E., & Varpio, L. (2024). Using framework analysis methods for qualitative research: AMEE Guide No. 164. *Medical Teacher, 46,* 603-610. https://doi.org/10.1080/0142159X.2023.2259073

Kuzle, A. (2018). Assessing metacognition of grade 2 and grade 4 students using an adaptation of multi-method interview approach during mathematics problem-solving. *Mathematics Education Research Journal, 30,* 185–207. https://doi.org/10.1007/s13394-017-0227-1

Livingston, J. A. (2003). Metacognition: An Overview. ERIC. https://eric.ed.gov/?id=ED474273

Lynch K., Hill H. C., Gonzalez K. E., Pollard C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, 41(3), 260–293.

Muijs, D., & Bokhove, C. (2020). *Metacognition and self-regulation: Evidence review*. Education Endowment Foundation.

Mulholland, K. (2021). Using pupil views to uncover evidence of children' s metacognition in mathematics. *International Journal of Student Voice*, *8*, 1‐35.

Mulholland, K., Gray, W., Counihan, C. & Nichol, D. (2025). Stepping Stones: Adopting a fading programme design to promote teachers' use of metacognitive strategies for mathematical problem solving, *Education Sciences, 15 (7),* 892. https://doi.org/10.3390/educsci15070892

Mulholland, K., Gray, W., & James, F. (2024). Early-stage programme development: Final report. Unpublished.

Muncer, G., Higham, P. A., Gosling, C. J., Cortese, S., Wood-Downie, H., & Hadwin, J. A. (2022). A meta-analysis investigating the association between metacognition and math performance in adolescence. *Educational Psychology Review*, *34*, 301‐334.

Ofsted (2023). *Coordinating mathematical success: The mathematics subject report.* Ofsted.

Ohtani, K., Hisasaka, T. (2018). Beyond intelligence: a meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacognition Learning* **13**, 179–212. https://doi.org/10.1007/s11409-018-9183-8

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. Contemporary Educational Psychology, 8(3), 293–316. https://doi.org/10.1016/0361-476X(83)90018-8

Perry, J., Lundie, D., & Golder, G. (2019). Metacognition in schools: what does the literature suggest about the effectiveness of teaching metacognition in schools? *Educational Review*, *71*(4), 483–500. https://doi.org/10.1080/00131911.2018.1441127

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), Issues in the measurement of metacognition (pp. 43–98). Lincoln: Buros Institute of Mental Measurements.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. Educational psychology review, 7(4), 351-371. https://doi.org/10.1007/BF02212307

Schraw, G. (2002). Promoting general metacognitive awareness. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 17–32). Champaign, IL: Springer.

Schraw & J. C. Impara (Eds.), Issues in the measurement of metacognition (pp. 43–98). Lincoln: Buros

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Van Herwegen, J., Anders, J. (2021). *What are the Characteristics of Teacher Professional Development that Increase Pupil Achievement? A systematic review and meta-analysis*. London: Education Endowment Foundation. The report is available from:

https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/teacherprofessional-development-characteristics

Son, L., Furlonge, N. and Agarwal, P. (2020). Metacognition: how to improve students' reflections on learning. Available at: Metacognition: How to Improve Students' Reflections on Learning (last accessed on 19 June 2025)

Treglia, M. (2018). *A comparison of offline and online measures of metacognition*. Senior theses, Trinity College, Hartford, CT. https://digitalrepository.trincoll.edu/theses/729

Xie, Y., Zeng, F., & Yang, Y. (2024). A meta-analysis of the relationship between metacognition and academic achievement in mathematics: From preschool to university. *Acta Psychologica*, *249*, 104486. https://doi.org/10.1016/j.actpsy.2024.104486

Veenman, M.V., Hout-Wolters, V., Bernadette, H.A.M. and Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations, *Metacognition and learning, 1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Vula, E., Avdyli, R., Berisha, V., Saqipi, B., & Elezi, S. (2017). The impact of metacognitive strategies and self-regulating processes of solving math word problems. *International Electronic Journal of Elementary Education*, *10*, 49‑59.

Wang, Y., Brown, C., & Dawson, J. (2024). *Developing Maths lesson planning and frameworks*: *Mastery, Logic and Reasoning in the Classroom*. Open University Press. 0335251803 · 9780335251803

Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R., & Ogbuehi, P. (2012*). Improving mathematical problem solving in grades 4 through 8: A practice guide (NCEE 2012-4055).* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http:// ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch/.