Peer-to-Peer Coaching
Evaluation Protocol

Evaluator (institution): RAND Europe
Principal investigator(s): Elena Rosa Speciani

Education Endowment Foundation

Template last updated: December 2022

Evaluation summary

Project title	**Using Peer-to-Peer Coaching to improve Maths attainment, a two-armed cluster randomised trial**
Developer *(Institution)*	CoachBright
Evaluator *(Institution)*	RAND Europe & University of Leeds (UoL)
Principal investigator(s)	Elena Rosa Speciani
Protocol author(s)	Elena Rosa Speciani, Ivana Cardamone, James Merewood, Erin Dysart, Louise Tracey
Trial design	Two-arm cluster randomised controlled trial with random allocation at school level
Trial type	Efficacy
Pupil age range and Key stage	11 – 12 (Year 7, KS3), 14 – 15 (Year 10, KS4)
Number of schools *(at design stage)*	100 schools
Number of pupils *(at design stage)*	30 per school – 15 from Year 7 and 15 from Year 10 (approximately = 3,000 pupils)
Primary outcome measure and source	Maths attainment (GL Progress Test in Maths (PtM) – Level 12 for Year 7; Level 15 for Year 10)
Secondary outcome measure and source	1. Sources of Mathematics Self-Efficacy Scale (SMSES) 2. Metacognition (Junior Metacognitive Awareness Inventory)

Protocol version history

Version	Date	Reason for revision
1.2 [*latest*]		
1.1		
1.0 [*original*]		N/A

Contents

Tables

Figures

Background and Objectives

Peer tutoring programmes have gained increasing attention within the educational policy landscape due to their potential to address persistent challenges in academic attainment, particularly in mathematics (Robinson et al., 2005; Roberts & Spangenberg 2020). Peer tutoring, as an evidence-based approach, aligns with national priorities to improve educational outcomes and reduce inequalities, particularly for Key Stage 3 (KS3) and Key Stage 4 (KS4) students. For example, the *Education Recovery Plan* launched in response to the COVID-19 pandemic emphasises the importance of targeted interventions to support pupils who have experienced learning loss. The plan includes funding for initiatives such as the National Tutoring Programme (NTP), which provides small-group and one-to-one tutoring for disadvantaged pupils (DfE, 2021). Peer tutoring complements this approach by offering scalable, cost-effective support that can be integrated into schools' broader recovery strategies, particularly for KS3 and KS4 students (Lynch et al., 2024)

The relevance of peer tutoring in mathematics is further amplified by the subject's foundational role in future academic and career opportunities. Mathematics proficiency is not only essential for success in STEM fields but also for broader economic participation (Pro Bono Economics 2021). The Department for Education (DfE) has emphasised the importance of improving maths attainment as part of its broader strategy to enhance social mobility and address skills shortages in the UK labour market (DfE, 2023). Peer tutoring programmes, which leverage the benefits of peer-led learning, offer a promising avenue to support these goals. They provide an opportunity to consolidate classroom learning, address misconceptions, and build confidence in pupils who may struggle with traditional teaching methods. In the context of mathematics, where underachievement and achievement gaps are well documented, peer coaching can be particularly important for providing targeted support, building confidence, and improving academic attainment (Cropp, 2017; Sharp, 2021).

Peer tutoring is identified as a highly effective approach to improving learning outcomes for both primary and secondary age pupils, with an average impact of five additional months of progress, based on extensive global evidence that includes mathematics tutoring (Alegre et al., 2025; EEF, 2021; Alegre et al, 2019a; Topping et al., 2004; Fisher, 2001). The positive effects of peer tutoring are seen to impact tutees and tutors equally (Tenenbaum, 2019; Roscoe & Chi, 2007; Topping, 2005; VanLehn et al., 2007). For tutees, tutoring approaches offer targeted, peer-led support to consolidate within class learnings, practice skills, and identify and overcome misconceptions (Topping et al., 2004; Yang et al., 2016). It has long been know that successful approaches can also help boost pupils' self-confidence and motivation for learning (Gardner, 1973). For the tutor, tutoring helps in revisiting and revising skills and prior knowledge as well as developing metacognitive understanding of topics (De Backer et al., 2016; Fatma & Engin, 2017; McKinstery & Topping, 2003).

Peer tutoring may be a useful intervention to target specific groups and could be especially beneficial in reducing the attainment gaps, such as those between disadvantaged pupils and their peers (DfE 2024). Disadvantaged pupils are considerably less likely to achieve higher grades at the end of secondary school. For example, in the 2022-2023 academic year, 25% of

disadvantaged pupils achieved grade 5 or above in their English and Maths GCSEs, compared with 52% of all other pupils (Social Mobility Commission, 2024). Although evidence specifically focusing on pupils from disadvantaged backgrounds is limited, research indicates that low-attaining pupils often gain additional benefits from peer tutoring, conceivably by providing targeted support to reinforce classroom learning, practise skills, and address misconceptions (EEF, 2021; Topping et al. 2011; Tymms et al., 2011; White, 2019).

Evidence in the literature also shows that peer tutoring can vary for pupils with SEND and EAL status (Topping, et, al, 2005). Similarly, research indicates that peer tutoring can enhance self-efficacy, with potential variations in impact based on gender dynamics within tutoring groups (Tenenbaum et al. 2019; Webb & Mastergeorge, 2004). While globally, peer tutoring is seen to have led to positive learning outcomes, evidence from the UK is more limited. However, an independent assessment of the CoachBright programme conducted by ImpactEd offers preliminary evidence of beneficial impacts on achievement, self-efficacy, and the enhancement of metacognitive skills (ImpactEd, 2023). However, the programme has not undergone a large-scale independent evaluation and will gain from a robust trial.

The current evaluation of the Peer-to-Peer Coaching programme represents a timely and important step in addressing this evidence gap. Designed to improve secondary mathematics outcomes, the programme targets disadvantaged pupils—a group disproportionately affected by attainment gaps. The intervention involves high-attaining Year 10 pupils tutoring Year 7 pupils who struggle with mathematics. Delivered by CoachBright, the programme includes weekly 60-minute sessions over a 10-week period, with Year 10 tutors receiving training to provide metacognitive guidance, to use tailored maths problems, and to understand mutual concept explanation. Year 10 tutors are selected based on their potential to achieve GCSE scores of 7-9 in mathematics, ensuring they possess the requisite skills and knowledge to support their peers effectively and low attaining Year 7 tutees are selected based on their Key Stage 2 (KS2) SATs results.

The evaluation is structured as a two-armed randomised controlled trial, with randomisation at the school level. At least 75% of pupils included in the trial will come from disadvantaged backgrounds, i.e., all Year 7 pupils and at least half of Year 10 pupils, ensuring the programme targets those most in need of support. The trial will measure the impact of the Peer-to-Peer Coaching programme on mathematics attainment, self-efficacy, and metacognition, providing robust evidence on its effectiveness. An embedded process evaluation will explore factors such as implementation fidelity, barriers and facilitators, and pupil responsiveness, offering valuable insights to inform future iterations of the programme. Following the efficacy trial, a pilot study during the 2026-2027 academic year will test a school-led model of implementation which will offer a differentiated package of support to schools to further their peer tutoring journey, test the programmes flexibility and scalability, and fit to different schools budgets. Additionally, it will further contribute to the evidence base. The aim of the pilot study will be to assess the feasibility of implementation when the school leads implementation of the programme with support from CoachBright. It will further assess the sustained impact of this less intensive model.

Intervention

CoachBright's Peer-to-Peer coaching programme aims to improve secondary mathematics outcomes through metacognitive guidance, tailored maths problems, and mutual concept explanation and is targeted at disadvantaged pupils. Specifically, high-attaining Year 10 coaches are provided training and support to coach disadvantaged Year 7 coachees who are lower attainers in meths, during 60-minute weekly sessions over the course of 10 weeks.

The core content of the peer coaching sessions centres around the younger pupils and their coaches developing and deploying metacognitive strategies in relation to mathematics. CoachBright offers two types of coaching: pastoral and academic. This evaluation will focus on academic coaching in mathematics. During the coaching sessions, peer coaches support the younger pupil to reflect on their academic progress and support them through specific topics they are finding hard using tailored mathematics problems, mutual concept explanation, and peer coaching for deeper understanding: coachees are encouraged to explain their reasoning to the peer coach, to promote a deeper understanding of mathematical concepts.

Further details of the programme, as designed to be implemented for this efficacy trial can be found in Table 1.

Table 1: Aspect of TIDieR[1]

Aspect of TIDieR	Exemplification relating to the evaluation
Brief name	Peer-to-Peer Coaching (P2P)
Why: Rationale, theory and/or goal of essential elements of the intervention	Evidence indicates that peer tutoring can significantly enhance both academic performance and social skills, especially when careful attention is given to selecting pairs and ensuring high-quality sessions (see EFF 2021a). While global research highlights promising outcomes, UK studies such as Paired Reading (Lloyd et al., 2014), Shared Maths (Lloyd et al., 2014), and PALS (Culora et al., 2022), have shown mixed results. RAND Europe and the University of Leeds are keen to evaluate CoachBright's Peer to Peer coaching (P2P) programme, as it provides an opportunity to expand the evidence base and explore previously unexamined aspects of peer tutoring in the UK. Indeed, peer tutoring is recognised as a promising approach in the EEF toolkit (see EFF 2021b), showing an average impact of +5 months.
Who: Recipients of the intervention	The P2P programme aims to benefit low-attaining Year 7 pupils and high attaining Year 10 pupils from disadvantaged backgrounds. *Maths attainment*

[1] http://www.bmj.com/content/348/bmj.g1687

	● Year 7 pupils with low attainment in mathematics (determined through their SATs results - SATs score < 100). ● Year 10 pupils with high attainment in mathematics (potential to achieve GCSE scores of 7-9 in maths, determined through one or a combination of end-of-year exams, teacher assessments, or any standardised tests taken in Year 9). *Disadvantage* ● At least 80% of the Year 7 pupils must be eligible for Free School Meals (FSM). Any other pupils identified as benefitting from the Peer-to-Peer coaching programme should meet one or more of the criteria for CoachBright's wider definition of disadvantaged (below). ● At least 50% of Year 10 pupils should be eligible for FSM or meet one or more of the wider disadvantage criteria (below), but their mathematics attainment should take precedence. We will not collect information on which criteria they meet, rather only whether they meet at least one criterion. Wider definition of 'disadvantaged pupils': ● Eligible for pupil premium funding (PP); ● Eligible for the Service Premium; ● Is a young carer; ● Is or has been a Looked After Child (LAC); ● Is known to be a refugee or asylum seeking child; ● Neither parent has attended higher education; ● Otherwise considered disadvantaged as identified by the school.
What: Procedures, activities and/or processes used in the intervention	The programme consists of an onboarding meeting with the School Coordinator and the Support Coordinator, a coach training session, a launch event, 10 coaching sessions and a graduation event at the end of the programme. There is also guidance for the Year 10 pupils throughout on completing their SSAT Leadership Accreditation. **School Coordinator onboarding meeting:** In this meeting with the School Coordinator and a member of the SLT, the CB Programme Manager provides the overview of the programme, sets expectations, confirms logistics, discusses pupil data required for the study and provides guidance around selecting and matching pupils. **Coach training:** The CB Programme Manager provides training to the 15 coaches. The coaches learn about

	CoachBright and the coaching journey (together with the 5 coaching tools of Active Listening, Effective Questioning, Building Trust and Relationships, Motivating Others and Giving Feedback). They also learn about session planning and how to gain and use coachee reflections. During the training, the coaches also practice planning a session. **Launch event:** In this event coaches and coachees get to know each other and discuss mathematical areas to focus on during the programme, so that the coaches can then use the information to plan the weekly sessions. The aim is to build relationships, set the goals of and expectations from the programme, and share logistics with the pupils. During the event, the School Coordinator and CoachBright Programme Manager observe the interaction among the pairs and if the pair does not seem to work, groups can be rearranged. **10 x weekly coaching sessions:** Each session begins with a 'starter' activity to make pupils comfortable and is then followed by the main activity. - Starter activity: this is used to settle the coachee into the session. It is often an activity that is fun and engaging and it is usually related to the main activity. For example, if the main activity is related to fractions, then the starter activity could be a fast-paced times table game as fractions understanding is underpinned by sound understanding of timetables. - Main activity: This activity is based on the gap(s) that the coach and coachee have discussed and identified to work on. It makes up the largest proportion of time in the coaching session. It is bespoke to the individual and can involve activities such as working on past exam questions, looking at a challenging homework problem, going back to basics on a mathematical topic that the coachee is finding complex, reinforcing concepts that aren't 'sticking' through activities and games. The main activity is planned by the coach and support is available from the school and CoachBright on how to do this effectively. The sessions end with reflections and planning for the following week. - CoachBright Programme Managers and School Coordinators encourage coaches and coachees to reflect in their coaching journals and share feedback with one another.

	Coaches After the coach training, the coaches reflect on what they have learnt, what they have enjoyed and what they are looking forward to. After each session the coaches reflect on which of the coaching tools they have used and also reflect on what went well and what could be improved in the following week. At the end of the programme the coaches are encouraged to reflect on what they are proud of and their coachee's progress. Coachees At the start of the programme the coachees are encouraged to reflect on their goals for the programme and are encouraged to break them down into smaller goals with actionable steps and a larger end goal. At the end of each coaching session the coachee is asked to reflect on what they learnt in the session, what went well, what could have been improved and what they would benefit from working on the following week. - CoachBright Programme Managers and School Coordinators hold a 10-15-minute one-to-one debriefs with coaches following each coaching session. During this time, they assist coaches in selecting mathematics resources from CoachBright google classroom, sites (e.g. Twinkl, TES) and in-school resources or in developing additional resources for the next session. The aim is to address specific challenges and knowledge gaps tailored to individual needs. **Graduation event:** At the end of the programme, coaches and coachees participate in a graduation event (often held at a local university), which is designed to celebrate their achievements and raise aspirations. The pupils receive certificates for their achievements. They will often have a tour of the university. Some universities offer workshops and Q&A sessions for the pupils or CoachBright will facilitate this if university staff or students are not available. The pupils are encouraged to reflect on their longer-term goals at the graduation as well as things they are proud of from the coaching programme. **SSAT accreditation for coaches:** At the end of the programme, coaches can earn a Schools, Students and Teachers Network coaching qualification. There is a specific

	support session midway through the programme to support the coaches with building their portfolio.
What: Physical or informational materials used in the intervention	The school will receive the necessary resources for the School Coordinators, coaches and coachees needed to implement P2P coaching programme. **School Coordinator onboarding meeting:** School Coordinator Handbook. **Coach training:** Training PowerPoint. **Launch event:** Launch PowerPoint and materials for games. **10 x weekly coaching sessions:** Coaches have access to different tools and resources which they will learn about during the coach training such as CoachBright google classroom, sites (e.g. Twinkl, TES) and in-school resources. **Graduation event:** Risk Assessment, Graduation PowerPoint (where applicable), Pupil certificates and goal materials. **SSAT accreditation for coaches:** SSAT criteria and support session.
Who: Intervention providers/implementers	CoachBright Programme Managers, School Coordinators ● Schools are required to identify a lead teacher (School Coordinator) who will be responsible for supporting with the delivery of the programme and a second contact (Support coordinator) who will act as the school coordinator in their absence. The person selected for these roles may differ between settings, depending on staff structure, capacity, and time to ensure pupils are well supported. ● The CB Programme Manager will provide an onboarding session and support to the School Coordinator to identify pupils that fit the eligibility criteria and match pupils to form 15 coaching pairs (15 Year 7 and 15 Year 10 pupils) as well as with facilitating the delivery of the coaching sessions and supporting the SSAT accreditation process. ● The School Coordinator will identify and match the pupils prior between the coach training and launch event. When pairing peer coaches and coachees, they are encouraged to match pupils with similar demographics, such as gender, or shared interests,

	like sports but schools know pupils best and can match pairs in the way that best suits the pupils. They will act as the main point of contact between CB and the school and help the programme manager to set-up and facilitate delivery. ● The CB Programme Manager will provide training to the Year 10 coaches. ● The CB Programme Manager and the School Coordinator will observe the matched pairs during the launch event and consider whether to rearrange pairings. ● The CB Programme Manager and the School Coordinator will facilitate the coaching sessions and the CB Programme Manager will administer the debrief sessions with the coaches at the end of each coaching session.
How: Mode of delivery	**The School Coordinator onboarding meeting:** This is a 1–2-hour long meeting delivered online or in person by the CB Programme Manager to the School Coordinator and a member of the SLT. **Coach training:** This is a 2–3-hour long group training delivered in-person by the CB Programme Manager to the 15 coaches selected in each school. It is typically delivered during the school day, with schools encouraged to use any existing PSHE, form tutor, or other pastoral time. **Launch event:** This is a 1–2-hour long event where coaches and coachees in each school meet in person. The CB Programme Manager facilitates this with the School Coordinators supporting. **10 x weekly coaching sessions:** These are in-person, one-to-one sessions where the peer coach works with the coachee. Each session is 50-60 minute long and take place during the school day, often during pastoral or PSHE periods. These sessions are usually scheduled at the same time and day each week and schools are advised to avoid scheduling these sessions during core subject times, such as English, Maths, and Science. **Graduation event:** This is an in-person event, usually at a local university. CB Programme Managers and university staff facilitate these events with the School Coordinators supporting by booking buses, sending consent forms, attending and supporting with activities on the day.

Where: Location of the intervention	The programme will focus on schools located in regions where CoachBright has established staff and school networks. Schools will initially be selected from the following regions: London (all London Boroughs), North East (all LAs), South East (Kent, Medway, Milton Keynes, Reading, Slough, Surrey, West Berkshire, West Sussex, Windsor and Maidenhead, Wokingham), South West (Cornwall and Devon), North West (Bolton, Bury, Manchester, Oldham, Rochdale, Salford, Stockport, Tameside, Trafford, Warrington, Wigan) West Midlands (all LAs), East Midlands (all LAs), East of England (Essex and Hertfordshire). However, depending on progress made during recruitment, and that a critical mass of roughly 10 schools is reached, schools may be considered for eligibility even when they are located outside of the above regions on a case-by-case basis.
When and how much: Duration and dosage of the intervention	10 x weekly, 50-60 minute long, in-person coaching sessions The study will take place across two cohorts – with the programme being rolled out to different schools in each of the two cohorts - to ensure that CoachBright can facilitate implementation effectively in a smaller number of schools at each point in time. ● Cohort 1 (C1): October 2025 – February 2026 ● Cohort 2 (C2): February – June 2026
Tailoring: Adaptation of the intervention	There are a few small adaptations which can happen during the course of the intervention: The main adaptation is switching of paired coach-coachees. At the launch event, the paired coach-coachee will be observed by the CoachBright Programme Manager and School Coordinator(s). If they feel that the pupils are not a good match (as seen when working with or observing pupils, through conversations with pupils or through session reflections or evaluations) then adaptions to the designated pairs is possible; i.e., rearranging the grouping during the early stages of the programme (up to week 3). In addition, if a coach is no longer available to conduct coaching sessions, then the coachee will join another coaching pair in a 2:1 ratio (2 coachees to 1 coach). Also, if a Year 7 coachee misses or fails to engage with three sessions schools can replace the coachee up to Week 4. The second adaptation is to the regions from which settings can be recruited into the trial. Ideally, settings will be recruited from the regions previous specified. However, on a

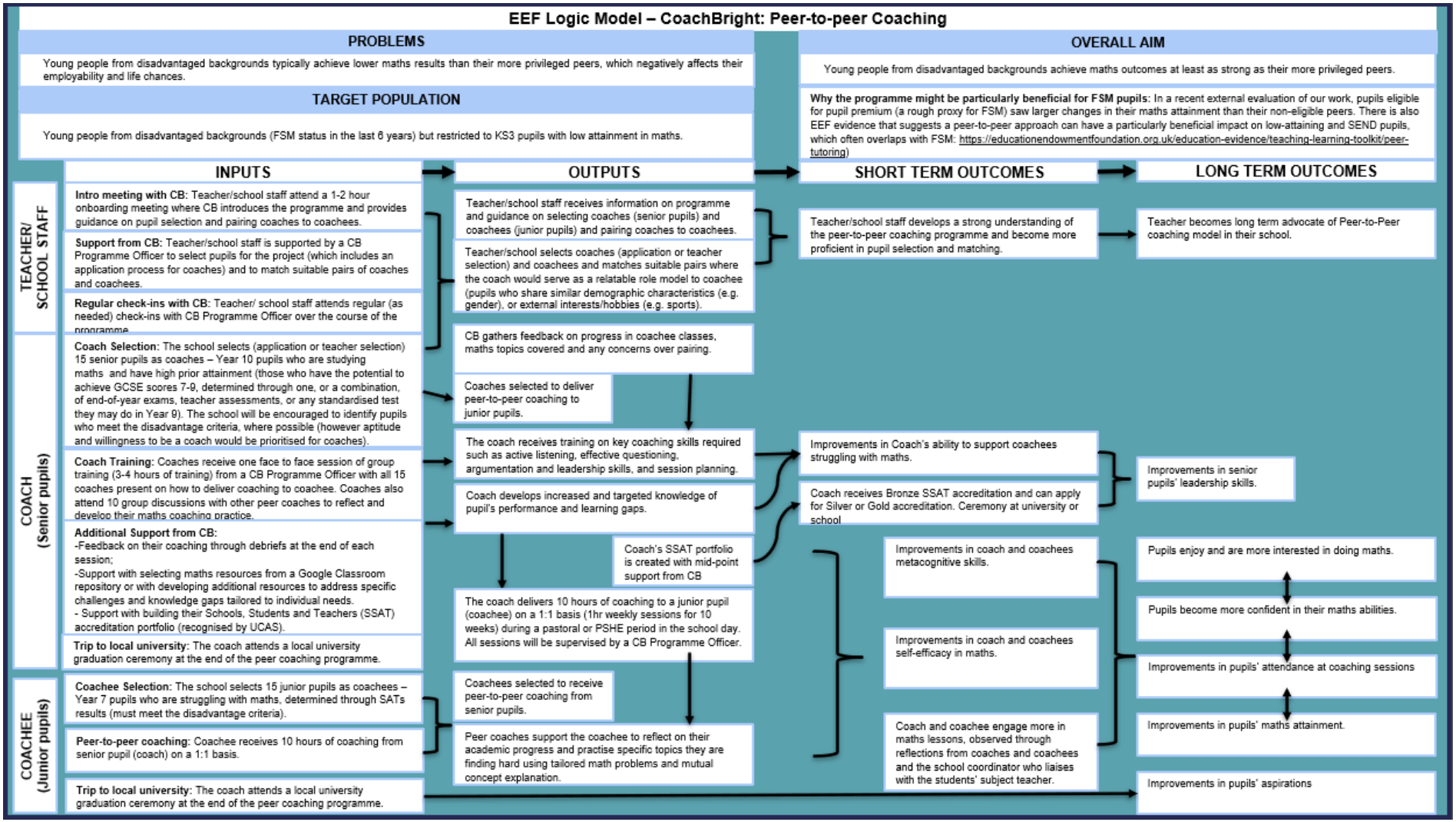
	case-by-case basis, settings from other regions may be considered for recruitment into the trial.
How well (planned): Strategies to maximise effective implementation	CoachBright support schools with pupil selection through a pupil selection document. Programme Managers are available via email or phone to clear up any clarifying questions. The School Coordinator will have an onboarding meeting with the Programme Manager where information about the programme will be given, and opportunities given for support and questions. Programme managers attend and facilitate each coaching session as well as facilitating the coach training, launch and graduation. Every session is supervised and adaptations to each session can be made according to what has been observed in previous weeks. Debrief sessions - coaches are encouraged to reflect on the session so that improvements can be made for the following week. This might include CB or the school providing a wider range of resources to support session planning. SSAT accreditation for coaches - the coaches are given all the relevant information about their SSAT accreditation during the coach training session. They are then given a SSAT session mid-way through the programme where the Programme Manager will look at the evidence the coaches have collected, make suggestions about future evidence to collect and can answer any clarifying questions that the school or coaches have.

The theory of change (ToC) underpinning the Peer-to-Peer Coaching programme can be found in Figure 1. Elements of the inputs and outputs are supported by research. For example, the support given to tutors in the Peer-to-Peer coaching programme is similar to existing evidence which suggests that to meet the expected high quality of peer interaction, pupils should be supported to tutor through training, questioning frames, and feedback for tutors (Fantuzzo et al., 1992; Fisher, 2001; Ginsburg-Block & Fantuzzo, 1997). Furthermore, the pairings between Year 7 and Year 10 align with research by Gaustad (1993) who suggested a gap of less than three years between the tutor and the tutee. This gap ensures that the tutor is confident with the topics covered while still providing a challenge for the tutee (Gustad 1993). Also, regular tutoring sessions for up to ten weeks have been found to be more effective than longer and less intensive programmes, specifically for outcomes related to understanding scientific concepts and vocabulary (Topping et al., 2004).

Evidence further supports the causal assumptions underpinning the intervention. For example, research suggests that peer tutoring improves mathematical reasoning and problem-solving skills by encouraging active engagement and dialogue between tutors and tutees- see Topping, (2005), and VanLehn et al. (2007). Research also supports the idea that peer tutoring programmes like Peer-to-Peer support self-efficacy (Roscoe & Chi, 2007) and metacognition (Webb and Mastergeorge., 2007). Furthermore, evidence suggests that metacognition and self-

efficacy could improve attainment in and of themselves, with the EEF identifying metacognition and self-regulation as high-impact strategies for improving attainment, with an average effect size equivalent to seven additional months of progress (EEF, 2021).

Figure 1: Logic model



Methods

Research Questions

Our primary research question is:

1. What is the impact of the CoachBright Peer-to-Peer (P2P) coaching intervention on mathematics attainment of Year 7 coachees, measured by the digital Progress Test in Maths (PtM), compared to similar pupils in control settings receiving business-as-usual?

Our secondary research questions are:

2. What is the impact of the CoachBright Peer-to-Peer (P2P) coaching intervention on maths attainment of Year 10 coaches, measured by the digital Progress Test in Maths (PtM), compared to similar pupils in control settings receiving business-as-usual?
3. What is the impact of the CoachBright Peer-to-Peer (P2P) coaching intervention on maths self-efficacy of Year 7 coachees, as measured by the Sources of Mathematics Self-Efficacy Scale (SMSES), compared to similar pupils in control settings receiving business-as-usual?
4. What is the impact of the CoachBright Peer-to-Peer (P2P) coaching intervention on metacognition of Year 7 coachees, measured by the Junior Metacognitive Awareness Inventory, compared to similar pupils in control settings receiving business-as-usual?
5. What is the impact of the CoachBright Peer-to-Peer (P2P) coaching intervention on maths self-efficacy of Year 10 coaches, as measured by the Sources of Mathematics Self-Efficacy Scale (SMSES), compared to similar pupils in control settings receiving business-as-usual?
6. What is the impact of the CoachBright Peer-to-Peer (P2P) coaching intervention on metacognition of Year 10 coaches, measured by the Junior Metacognitive Awareness Inventory, compared to similar pupils in control settings receiving business-as-usual?

Trial design

Table 1: Trial design

Trial design, including number of arms		Two-arm, cluster randomised controlled trial
Unit of randomisation		School
Stratification variables (if applicable)		Region: The region or LA in which the school is located. Setting type: local authority, multi academy trust, a free school or a grammar school.
Primary outcome	**Variable**	Mathematics attainment
	Measure (instrument, scale, source)	Digital Progress Test in Maths (PtM) age standardised score 69 – 141, GL Assessments Maths attainment (GL Progress Test in Maths (PtM) – Level 12 for Year 7; Level 15 for Year 10)

Secondary outcome(s)	**Variable(s)**	1. Maths self-efficacy 2. Metacognition
	Measure(s) (instrument, scale, source)	1. Sources of Mathematics Self-Efficacy Scale (SMSES), 24 - 120, Usher & Pajares 2009 2. Junior metacognitive awareness inventory, 1 – 18, Sperling et. al, 2002.
Baseline for primary outcome	**Variable**	KS2 SATS maths attainment
	Measure (instrument, scale, source)	KS2 maths scores (KS2_MATSCORE), 0 – 999, acquired from the NPD
Baseline for secondary outcome	**Variable**	KS2 SATS maths attainment
	Measure (instrument, scale, source)	KS2 maths scores (KS2_MATSCORE), 0 – 999, acquired from the NPD

The efficacy study is designed as a two-armed randomised controlled trial with the unit of randomisation at the school level and the unit of analysis at the pupil level. The treatment arm will receive Peer-to-Peer programme at a reduced cost, while the control condition will continue with business as usual and will receive £750 in compensation for their support with the trial and associated data collection activities. The primary outcome is mathematics attainment with secondary outcomes being metacognition and maths self-efficacy.

Participants

Schools

Recruitment will be led by CoachBright supported by EEF, RAND Europe, and University of Leeds. Schools will be recruited during the academic year 2024-25. Schools will be recruited from the eight geographical areas of London (all London Boroughs), North East (all LAs), South East (Kent, Medway, Milton Keynes, Reading, Slough, Surrey, West Berkshire, West Sussex, Windsor and Maidenhead, Wokingham), South West (Cornwall and Devon), North West (Bolton, Bury. Manchester, Oldham, Rochdale, Salford, Stockport, Tameside, Trafford, Warrington, Wigan), West Midlands (all LAs), East Midlands (all LAs), East of England (Essex and Hertfordshire). Other locations may be considered on a case-by-case basis with key decisions made on how feasible it will be for CoachBright to deliver the intervention with fidelity in the other locations.

Recruitment methods for the trial will include information webinars, 1:1 calls, a video explaining the trial, mass mail outs via email, social media posts, known contacts and attending education events.

Schools will be regarded as recruited providing they sign the MoU and the Data Sharing Agreement, and they meet the following eligibility criteria:

- They are a local authority, multi academy trust, a free school or a grammar school.
- They will have pupils in Year 7 and 10 on the same site during the 2025-2026 school year.

- They are from the following regions and Local Authorities (LAs), although other locations will be considered on a case-by-case basis: London (all London Boroughs), North East (all LAs), South East (Kent, Medway, Milton Keynes, Reading, Slough, Surrey, West Berkshire, West Sussex, Windsor and Maidenhead, Wokingham), South West (Cornwall and Devon), North West (Bolton, Bury. Manchester, Oldham, Rochdale, Salford, Stockport, Tameside, Trafford, Warrington, Wigan) West Midlands (all LAs), East Midlands (all LAs), East of England (Essex and Hertfordshire).
- They will be able to identify 15 coaching pairs (15 Year 7 and 15 Year 10 pupils) that fit the pupil eligibility criteria (see further information on pupils below).

Schools will <u>not</u> be eligible for the trial if any of the following apply:

- They are a private school or special school.
- They will be participating in another mathematics-focused EEF trial for the same year groups in 2025-26 such as Action Tutoring, Fluent and Flexible Calculators or Specialist Knowledge for Teaching Mathematics.
- They are currently delivering a CoachBright Peer-to-Peer coaching programme.

Schools will be required to identify a lead teacher (School Coordinator) who will be responsible for identifying 15 coaching pairs (15 Year 7 and 15 Year 10 pupils) that fit the eligibility criteria and support with the delivery of the programme in intervention schools. Schools will also be required to identify a second contact (Support Coordinator) who will act as the School Coordinator in their absence. The Support Coordinator is usually a member of the SLT.

Pupils

Pupils in recruited schools will be eligible for inclusion in the research if:

Maths attainment

- They are Year 7 pupils with low attainment in maths (determined through their SATs results - SATs score < 100)
- They are Year 10 pupils with high attainment in maths (potential to achieve GCSE scores of 7-9 in maths, determined through one or a combination of end-of-year exams, teacher assessments, or any standardised tests taken in Year 9)

Disadvantage

- At least 80% (12 out of 15) of Year 7 pupils should be eligible for Free School Meals (FSM). Any other pupils identified as potentially benefitting from the Peer-to-Peer coaching programme should meet one or more of the criteria for CoachBright's wider definition of disadvantaged (below).
- At least 50% (8 out of 15) of the Year 10 pupils should be eligible for FSM or should meet wider disadvantage criteria but their maths attainment should take precedence.

Wider definition of 'disadvantaged pupils':

- Eligible for pupil premium funding (PP);

- Eligible for the Service Premium;
- Is a young carer;
- Is or has been a Looked After Child (LAC);
- Is known to be a refugee or asylum-seeking child;
- Neither parent has attended higher education;
- Otherwise considered disadvantaged as identified by the school.

Parents or carers are provided with a privacy notice and project information sheet, along with a withdrawal form. Parents have two options to withdraw from the evaluation. Firstly, via the withdrawal form at the start, and then via email to the evaluation team during the evaluation. We do this to ensure that parents can opt out at any time during the period of time that the evaluation team holds their child's data. There will be separate, opt-in consent forms for interviews, surveys, and focus groups (reviewed and approved by UoL ethics committee). This includes opt-in consent for pupils who take part in participant focus groups.

A pupil can still take part in the intervention if their parent withdraws their child from their data being used in the study.

Incentives

Schools assigned to the intervention group will be required to pay a subsidised fee of £500 per school in order to receive the programme (the standard cost is £170 per pupil, meaning the full costs outside this evaluation would be a total of £5,100 for 30 pupils).

Schools assigned to the control group (business as usual) will receive £750 as a 'thank you' (with £250 paid after initial pupil data collection, and the remaining £500 to be paid at the end of the trial, following completion of all the trial requirements). This is intended to compensate for the time and effort spent participating in activities related to the evaluation. Control schools will be offered the opportunity to access one of ten possible places on the school-led model of the programme after the trial in the 2026-2027 school year free of cost.

Outcomes measures

Baseline measures

The baseline measure for maths performance will be the KS2 maths scores from the NPD, specifically the KS2_MATSCORE variable which is available for all children from 2015 onwards. This measure was chosen as it is a valid and reliable measure, a statutory requirement of schools in England delivered under exam conditions when pupils are in Year 6 and marked externally by the DfE. This approach also reduces the burden on schools and is in line with EEF guidance (see EEF, 2022a).

This measure is available for both year groups, however we will need special permission for Year 10 as establishments' IDs (school name, URNs, LAESTABs, etc.) together with any attainment data from 2019/20 or 2020/21 is restricted as the Department for Education (DfE) does not wish to hold schools accountable for attainment during the COVID-19 pandemic. A baseline allows us to improve pre-post correlation estimates, which in turn will improve power with a smaller

number of settings and also allows for exploration of differential attrition, if necessary (see EEF 2022).

Primary outcome

The primary outcome of the study will be maths attainment. GL's Progress Test in Maths (PtM) will be used to measure maths attainment, as it is a valid and reliable measure known by schools, and easy to administer. We will use a digital form of PtM (PtM12 and PtM15) for the following reasons. Firstly, it is suitable for use with both Year 7 (PtM12) and Year 10 (PtM15) pupils. Additionally, the PtM15 is only digital and it is computer adaptive, which guards against floor or ceiling effects. One limitation is that PtM15 has not been standardised with a UK population. However, PtM has been previously used in several EEF trials, demonstrating robust properties, with no floor or ceiling effects (Demack, et al., 2022).

PtM measures mathematical content knowledge (curriculum content category) and 'Understanding and applying mathematical processes through reasoning and problem solving' (process category). Age standardised and raw scores are provided as well as number of questions attempted, percentage accuracy broken down by question, by curriculum content area or by process category. This assessment takes 45-75 minutes to complete (depending on the level).

The PtM will be administered by University of Leeds, who will send in independent testers, that they have trained, to oversee the testing. The test will require the use of computers in schools, however there is budget available for a set of laptops. Upon test completion, results are generated automatically by GL assessment's system, ensuring that scoring variability across administrators will be minimised. Raw and standardised scores will be provided by GL assessment.

In selecting a primary measure for this evaluation, we considered the differential impacts the programme may have on different year groups, wanting to ensure that we selected a measure that could encompass all attainment levels. The impact on Year 7 pupils is likely to be more pronounced than on Year 10 pupils, as Year 10s are chosen based on their inherently higher attainment levels.

Endline assessments will take place at the end of the intervention. This will be January – February 2026 for Cohort 1 and May – June 2026 for Cohort 2.

Secondary outcome

Maths self-efficacy

Given the research evidence of self-efficacy's role in mediating academic attainment (Wang et al., 2024), its influence on long-term achievement in maths (Parker et al., 2013), and the presence of maths self-efficacy as a short-term outcome on Peer-to-Peer's Theory of Change (ToC), we have included maths self-efficacy as a secondary outcome.

Mathematics self-efficacy should be measured using a maths-specific tool because it is a domain-specific construct that more accurately predicts performance and informs targeted

educational interventions (Bandura, 1997). This approach ensures the assessment is relevant to the mathematical tasks and challenges, providing insights that general self-efficacy measures cannot offer (Zimmerman, 2000; Usher & Pajares, 2008).

We will use the Sources of Mathematics Self-Efficacy Scale (SMSES) (Usher & Pajares 2009) which is specifically tailored to self-efficacy behaviour in maths. This will be administered by University of Leeds to Year 7 and Year 10 pupils at the end of the intervention (Cohort 1 February/March 2026, Cohort 2 May/June 2025). Pupils will be randomly allocated to sit one of the SMSES or metacognition tests (see below) to avoid burden.

The SMSES was refined and validated through a multi-phase study by Usher and Pajares (2009), involving middle school students. The final version of the SMSES consists of 24 items across four subscales, each corresponding to one of Bandura's sources. The items are designed to capture students' perceptions of their experiences and feelings related to mathematics learning, making the scale particularly useful for educational research and intervention evaluation. Confirmatory factor analysis supported the four-factor structure, and the model was found to be invariant across gender and ethnicity, indicating its robustness across diverse student populations. The subscales showed significant correlations with related constructs such as mathematics self-concept, mastery goal orientation, and optimism, supporting the scale's convergent validity. These properties make the SMSES a reliable and theoretically grounded tool for measuring the underlying sources of students' mathematics self-efficacy.

The SMSES was developed for use with US middle school pupils (ages 11 – 14). However, a systematic review of interventions aimed at improving mathematics self-efficacy note that the SMSES has been used in studies exploring outcomes in secondary and post-secondary settings (Liu et al., 2022). The SMSES was developed in the US but has been used in a number of other countries (Liu et al., 2022). Despite evidence that the SMSES is valid and reliable outside the US, and with older age pupils, we will be conducting a pilot in three schools before endline to check that the measure is valid and reliable with our chosen population. If this measure is not valid, it will not be used, and alternatives will be sought. Results of this will be presented in the final report as an appendix.

Metacognition

Additionally, we will explore the impact on metacognition. Despite strong evidence overall to suggest that metacognitive knowledge directly impacts maths outcomes (Donker, et. al, 2014), few trials in the UK have explicitly established this link. In particular, as a funder, EEF's justification for funding this project was funding priority 1a: Mathematical reasoning: approaches designed to develop learners' cognitive, metacognitive and self-regulative knowledge and skills, specifically in maths.

We will use the Junior Metacognitive Awareness Inventory (JMAI) (Sperling et al., 2002) designed to assess children's metacognition. The JMAI has been rated as suitable for use in both KS3 and KS4 by the EEF.[2] It consists of 18 Likert style questions which measure components of pupils'

[2] https://educationendowmentfoundation.org.uk/measures-database/junior-metacognitive-awareness-inventory-age-11-15

metacognition, namely knowledge of cognition and regulation of cognition. Pupils score themselves from 1 to 5, with higher scores indicating higher knowledge and regulation of cognition. Research by Sperling et al., (2012) demonstrates that the JMAI effectively assesses metacognitive awareness in adolescents, showing high internal consistency and construct validity, which supports its use in educational settings to evaluate students' metacognitive skills (Sperling et al., 2012). Additionally, the inventory's reliability is evidenced by consistent results across diverse populations, making it a dependable tool for educators and researchers alike (Sperling et al., 2012).

Secondary outcome administration

Pupils participating in the evaluation will be randomly selected to receive either the SMSES or the JMAI, both of which take a few minutes to complete, but they will not be asked to complete both to reduce test burden. Both of these will be paper-based tests. Pupils in each Year group in each school will be randomly assigned to complete one of the two tests, resulting in an approximately half of pupils in Year 7 completing the MSES and half of pupils in Year 7 completing the JMAI – and the same for Year 10. RAND Europe will provide University of Leeds with randomised pupil lists which administrators will use to assign pupils to either SMSES or JMAI. University of Leeds will train and supervise independent qualified and trained testers to administer these.

Sample size

Table 2: Sample size calculations for Year 7

		Overall	FSM
Minimum Detectable Effect Size (MDES)		0.151	0.165
Pre-test/ post-test correlations	level 1 (pupil)	0.75	0.71
	level 2 (class)	n/a	n/a
	level 3 (school)	0.61	0.54
Intracluster correlations (ICCs)	level 2 (class)	n/a	n/a
	level 3 (school)	0.07	0.07
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	
Average cluster size		15	13
Number of schools	Intervention	50	50
	Control	50	50
	Total	100	100
Number of pupils	Intervention	15	13
	Control	15	13

	Total	30	26

The minimum detectable effect size (MDES) for this study has been calculated using a two-level random assignment design, to reflect the design of the trial, with randomisation occurring at the school level and analysis occurring at the individual level. We have calculated MDES for Year 7 and Year 10, for all pupils, and for the FSM subset. In calculating the MDES for all groups, we have made a number of assumptions: randomisation at the school level with 50:50 allocation, alpha of 0.05 and power at 0.8. All MDES calculations were made using PowerUp!.

Additionally, for Year 7 (overall, see Table above), we assumed an intra-cluster correlation of 0.10[3] and a pre-test/post-test correlation of 0.75 at the pupil level and 0.61 at the school level, with 15 pupils per cluster.[4] These assumptions gave us a MDES of 0.151. As is standard in EEF trials, we will run a subgroup analysis on children from disadvantaged backgrounds, using the FSM status[5]. Accounting for this, changing the pre-test/post-test correlations to 0.71 at pupil level and 0.54 at school level,[6] and adjusting the average cluster size to 13 pupils (as we are assuming at least 80% will come from disadvantaged backgrounds), we found a MDES of 0.165 for Year 7 pupils with FSM status.

We also computed the MDES for Year 10 sample, overall and FSM pupils. We assumed an intra-cluster correlation of 0.10 and a pre-test/post-test correlation of 0.73 at the pupil level and 0.53 at the school level, based on a recent paper by the EEF (Singh et al., 2023).[7] These assumptions gave us a MDES of 0.183. For the FSM subgroup analysis, we assumed an intra-cluster correlation of 0.05, a pre-test/post-test correlation of 0.67 at pupil level and 0.45 at school level, based on a recent paper by the EEF (Singh et. al, 2023).[8] We adjusted the average cluster size to 8 pupils to reflect the fact that 50% of the Year 10 sample will be composed of FSM pupils. These assumptions gave us a MDES of 0.219 for FSM pupils.

MDES assumptions with attrition

Finally, we also explored the impact of average pupil and school level attrition on the MDES using the same assumptions above but using attrition at 26% pupil-level and 12% at school level.[9] The results are listed in the table below (Table 3).

Table 3: Attrition

	School-level attrition (12%)	Pupil-level attrition (26%)	Pupil and school level attrition
Year 7 (overall)	0.171	0.171	0.183

[3] Based on previous EEF trials of Year 7 maths interventions, using KS2 at baseline and GL PtM at endline.
[4] Based on previous EEF trials we calculated level 1 and level 2 correlations for our proposed outcomes (KS2 SATs and Progress Test in Maths (PtM)) for pupils in Year 7.
[5] Using the EVER6_FSM_P variable from the NPD
[6] Based on a review of data from published EEF secondary school maths trials.
[7] Using the estimates from the EEF model for KS4 maths.
[8] Using the fixed estimates from the EEF model for FSM subgroup analysis in KS4 maths.
[9] Based on attrition reported in published EEF secondary maths trials: The Rise Project, Maths in Context, Fit to Study, Realistic Maths Education, and Mathematics Mastery.

In considering the ideal sample size we took several factors into consideration. Firstly, the average effect size from secondary peer tutoring trials taken from the EEF toolkit is 0.36 (EEF, 2021), and is similar to the effect size of 0.38 found in a meta-study of secondary school peer maths tutoring studies (Alegre et al., 2019b)[10]. However, effect sizes in EEF trials of maths interventions in secondary schools have tended to be much smaller – between 0.01 and 0.03 (Jerrim et al. 2015; Husain et al., 2019; Demack et al. 2022), suggesting the need to be conservative with sample size calculations.

Randomisation

In designing the trial, we have carefully considered several factors and have chosen school-level randomisation for a variety of reasons.

Firstly, there may be a lack of eligible Year 10 pupils with individual-level randomisation. To maximise CoachBright's offering, 15 pairs per school are needed. With a 50:50 allocation in individual randomisation, schools would need to identify 30 eligible disadvantaged and high-performing Year 10 pupils, which could be challenging, even if non-disadvantaged pupils are included.

Additionally, school-level randomisation is likely to result in reduced post-allocation demoralisation, as noted in previous EEF trials (Demack et al., 2016). While a waitlisted design could mitigate this in individual randomisation, it may not fully address initial demoralisation, and schools might struggle to accommodate two tutor sets in one year.

Moreover, there is a significant risk of contamination with individual randomisation. This approach would mean that each school contains both treatment and control group pupils. Such a setup could lead to situations where pupils receiving the treatment share skills or resources with those in the control group, especially if they have friends or siblings in other year groups who could benefit from the intervention. These scenarios could result in inaccurate impact estimates due to bias from control group members receiving a modified form of the intervention. Therefore, randomisation will be implemented at the school level.

Randomisation will take place after baseline data collection in October 2025. Schools will be randomly assigned either to receive the Peer-to-Peer coaching programme (the intervention group) or to carry on with teaching and learning as normal (the control group). The sample will be evenly split between the treatment and control groups. The randomisation code will be detailed fully in the Statistical Analysis Plan (SAP) due to be published in January 2026.

Randomisation will be conducted by RAND researchers who will be blind to allocation (i.e., by assigning schools meaningless identifiers by a separate member of the research team). Randomisation will be stratified by region, with settings as the unit of randomisation and pupils the unit of analysis. This randomisation approach is driven by the school-level approach of the intervention, which requires that pupils in each school that are involved in the program are assigned to the same condition. Once randomisation is complete, RAND Europe will inform

[10] We also note that not all studies in this meta-analysis are from Randomised Controlled Trails (RCTs) which may overinflate effects.

CoachBright about the treatment allocation of settings. CoachBright will in turn assign half of the treatment group to cohort 1, and half of the treatment group to cohort 2; this decision will be made based on resource allocation and the feasibility of coordinators to travel between settings to deliver the programme. RAND Europe will randomly assign half of the control group to cohort 1 and half of the control group to cohort 2.

Statistical Analysis

The statistical analysis will follow EEF statistical guidance (2022) and will be described in detail in a SAP which will be prepared within three months of randomisation (January 2026). The proposed analysis is provided in brief below.

Analyses will be conducted on an intention-to-treat (ITT) basis, following an analysed-as-randomised approach.

Primary analysis

The primary analysis will investigate any difference in maths attainment (GL Progress test scores) between treatment and control schools. The raw scores will be obtained for the outcome measure, as well as a Standardised Age Score (SAS) where this is generated on sample which is representative of UK schools.[11]

A linear mixed effects regression model will be used to estimate the adjusted mean difference in scores, with analysis undertaken at the pupil-level. Year 7 and Year 10 pupils will be analysed separately to understand the differential effects on each group in isolation, facilitating the identification of specific impacts for each year group. This separate analysis also enhances statistical precision and power by reducing variability (Bryk & Raudenbush, 1992).

Additionally, we will explore school-level effects through an exploratory analysis, accounting for confounding variables (Ray et al., 2022). This will be done either by pooling the Year 7 and Year 10 cohorts or analysing them separately and then aggregating effect sizes through meta-analysis. We will select the most appropriate approach and use the alternative method as a robustness check to ensure the reliability of our findings.

The main analysis consists of the model for outcomes of pupils nested in settings, which is:

(1) $Y_{ij} = \beta_0 + P2P_j\tau + Z_j\beta_1 + X_{ij}\beta_2 + u_j + e_{ij}$

where Y_{ij} is the GL Progress Test in Maths score for child $_i$ in school $_j$; β_0 is the cluster-level coefficient for the slope of a predictor on number skills; $P2P_j$ is a binary variable (a value of 0 or 1) which indicates whether the school was assignment to receive the intervention [1] or was assigned to the control group [0]; Z_j are school-level characteristics, here the stratifying variable

[11] GL assessments PtM provides an age standardised test score for Year 7 cohorts, and for Year 10 cohorts where they take undertake the PtM14 assessment, but not where they undertake the PtM15 assessment. We believe that Year 10 cohorts may need to undertake the PtM15 assessment, depending on when the endline assessment is completed, and therefore may collect only raw scores.

of geographical location (if used for randomisation); X_{ij} represents characteristics at child level (child$_i$ in setting $_j$), specifically the KS2 maths score which is used here as a baseline outcome; u_j are setting-level residuals and e_{ij} are individual-level residuals.

Equation (1) is specified as a 'random intercepts' model. Since $\beta_{0j}=\beta_0+u_j$, the setting-specific intercept for school *j* is random by design (i.e., it is a number which could take any value); further, the assumed distribution is given as $\beta_{0j} \sim_{i.i.d} N(\beta_0, \sigma_u 2)$. We are interested in a target parameter (i.e. the focal result of the trial) is τ, a binary treatment/control indicator variable. An effect size (*Hedge's G*) will be standardised using unconditional variance in the denominator. In addition, in line with EEF guidance (see EEF 2022a), we will report 95% confidence intervals in order to communicate any statistical uncertainty. This will tell us the average effect of the intervention on numeracy outcomes in treatment settings (i.e., those that receive the intervention) compared to those in control settings (i.e., those that did not).

Secondary analysis

The secondary outcomes – (1) Sources of Mathematics Self-Efficacy Scale (SMSES) scores and (2) Junior Metacognitive Awareness Inventory scores – will be analysed using the same specification as in Equation (1) listed under the 'Primary outcome analysis' above, however, we will substitute the PtM scores (i.e., Y in Equation (1)) for the secondary outcome measure scores. Since both measures produce different scales/scores it is likely that analysis will be carried out on each secondary outcome measure individually, rather than through the use of factor variables generated as a proxy for an endline performance in a self-efficacy measure.

Sub-group analysis

For sub-group analyses, we propose to collect FSM, SEND, EAL and gender data from the National Pupil Database (NPD), along with the KS2 maths scores needed for the primary analysis. The NPD variables to identify these pupils in will be: EVERFSM_6_P, SENprovision and LanguageGroupMajor. The analyses of the subgroup will use two approaches, as suggested in EEF analysis guidance (EEF, 2022a).

Free school meals sub-group analysis

The trial's main focus is on disadvantaged students, thus FSM sub-group analysis will be conducted. Unique Pupil Numbers (UPNs) will be utilized to retrieve further information about pupil characteristics from the NPD. The impact of the intervention on pupils eligible for FSM will be examined by incorporating FSM status (the EVERFSM indicator, EVERFSM_6_P in the NPD) and an interaction term between FSM status and allocation in the primary analysis model. Furthermore, the primary analysis will be conducted again specifically for the subgroup of FSM pupils within each year group separately (assuming the sample size of FSM pupils allows for this).

The first analysis will run the primary model given in Equation (1) on the FSM subgroup only. Effect sizes and statistical uncertainty will be calculated on the FSM subgroup following the procedure outlined in the above section on *Primary Analysis*.

In order to make use of the entire sample, the treatment effect on the FSM subgroup will also be estimated using an interaction model:

$$(2)\ \boldsymbol{Y_{ij} = \beta_0 + P2P_j\tau + FSM_j\beta_1 + (FSM_j * P2P_j)\beta_2 + Z_j\beta_3 + X_{ij}\beta_4 + u_j + e_{ij}}$$

This is the same model specification as in equation (1), with the addition of the FSM_j indicator of disadvantage and an interaction term combining FSM eligibility and treatment allocation $FSM_j * P2P_j$).

The primary coefficient of interest in the interaction model is β_2, which can be interpreted as the additional treatment effect experienced by children with FSM conditions: a positive β_2 is indicative of a treatment acting as a 'gap-closer' and a negative β_2 indicative of treatment acting as a 'gap-widener'. The treatment effect size will be calculated by hand using the coefficients in the interaction models and the unconditional standard deviation of the FSM sub-sample[12], according to EEF guidance (2022), and compared with that calculated from the model on the FSM sub-sample. In accordance with EEF guidelines, and with the procedure outlined above for FSM, we will again generate two estimates of effect size of the treatment on the sub-sample: i) by running the primary model given in equation (1) on the FSM sub-sample, and ii) by running the analysis on the entire sample using the interaction model specified in equation (2) and calculating the treatment effect size by hand.

SEND, EAL status and Gender sub-group analysis

For the pupils with special education needs or disability, we will use the variable SENprovision from the NPD as it is the most relevant measure to the target population. The analysis will be then undertaken with the binary SEND variable (SEND - eligible=1; non-SEND-eligible=0).

$$(3)\ \boldsymbol{Y_{ij} = \beta_0 + P2P_j\tau + SEND_j\beta_1 + (SEND_j * P2P_j)\beta_2 + Z_j\beta_3 + X_{ij}\beta_4 + u_j + e_{ij}}$$

Similarly, for sub-group analysis on EAL status and gender, we will use the variable LanguageGroupMajor from the NPD to investigate the EAL status and the gender variable to investigate the effect on gender. We will conduct the same analysis described above using the following interaction models (for EAL status and gender, respectively):

$$(4)\ \boldsymbol{Y_{ij} = \beta_0 + P2P_j\tau + EAL_j\beta_1 + (EAL_j * P2P_j)\beta_2 + Z_j\beta_3 + X_{ij}\beta_4 + u_j + e_{ij}}$$

$$(5)\ \boldsymbol{Y_{ij} = \beta_0 + P2P_j\tau + GENDER_j\beta_1 + (GENDER_j * P2P_j)\beta_2 + Z_j\beta_3 + X_{ij}\beta_4 + u_j + e_{ij}}$$

Analysis in the presence of non-compliance

By design, the ITT approach considers whether it was intended for each individual to receive the intervention, rather than whether that individual actually received the intervention in practice. Therefore, we will also examine the effect of the treatment on the primary outcome, specifically for those who complied with the intended programme of the intervention.

[12] This is calculated according to the following formula: $\mathrm{P2Pj}\tau + \mathrm{P2Pj}\tau * \mathrm{SENDi2sd}$

As the ITT approach is inherently conservative, our compliance analysis will also explore the impact of compliance, which will examine whether differences exist depending on the number of session individuals attend. Coach and coachee session attendance logs will be collected by CoachBright (i.e., through an attendance register) and shared with RAND Europe and the University of Leeds before analysis takes place.

Compliance was discussed by the EEF, CoachBright and the Evaluation team and has been defined as: at least 70% pupil attendance at coach training and 70% of weekly coaching sessions. This measure of compliance is therefore defined at the pupil level using a continuous variable and will form part of a two-stage-least-squares approach that retains the logic used during the randomisation process.

In a situation of imperfect compliance, whereby not all intervention schools are deemed compliant according to the above criteria, we will undertake a complier average causal effect (CACE) analysis, using two-stage least squares (2SLS) estimation to recover the local average treatment effect (LATE) of attending the coaching sessions on the primary outcome. Further details of how compliance will be analysed will be presented in the SAP.

Additional analyses and robustness checks

Imbalance at Baseline

If randomisation is conducted as intended, it will yield groups which are equivalent at baseline, with any imbalance at baseline occurring by chance. However, we will conduct a check for imbalance at baseline in the realised randomisation; baseline equivalence testing will be conducted at the school and pupil level. At both the school and pupil level, we will check the balance by means of cross-tabulations and histograms over specific characteristics, including baseline assessment and the proportion of the sample in each school who are from a disadvantaged background.

Exploratory analysis

Research indicates that the gender of peers in secondary schools significantly influences outcomes (Smith & Andersen, 2022) and progression into STEM subjects (Riegle-Crumb, et al., 2006), particularly for girls when their female peers also have an interest in STEM (Raabe, et al., 2019). To contribute to this important body of literature, we propose that a key aspect of this study could be examining the differential impact of same-gender and mixed-gender pairs in Peer-to-Peer learning. Schools will provide the University of Leeds (UoL) with gender information, along with students' names, dates of birth, year groups, and Unique Pupil Numbers (UPN). UoL will then share this information with RAND Europe. Assuming that there is sufficient variation in the genders of mentor/mentee pairs, we will aim to conduct exploratory analyses on the differences between same-gender and mixed-gender pairs of coaches and coachees. This approach will help us understand the nuanced effects of gender dynamics in peer tutoring settings.

Furthermore, literature suggests that the ability of peers can influence outcomes in secondary schools. Therefore, we will examine whether there is natural variation in Year 10 coach maths outcomes at baseline, in order to provide an understanding on whether being matched with a higher ability Year 10 coach has a larger impact on Year 7 coachee outcomes. This analysis may not be possible upon further examination of baseline maths scores among the Year 10 coaches; i.e., it may not be possible to draw sufficient distinction between ability if all Year 10 pupils are of similar ability.

Missing data analysis

Missing data can arise through different routes, whether through non-response (a school does not provide some information about a participant) or attrition (a participant removes themselves from the sample through opting-out of the evaluation). This can occur at both the school and pupil level. While it is important to include as much data as possible (in order to preserve the intended sample), it can be problematic to apply the ITT analysis approach when we are not able to complete endline testing for all school/pupils which were randomised. In order to understand the extent and pattern of missingness in data, and whether this has an impact on the analysis, we will explore missingness further.

Initially, we will explore the level of attrition across both the treatment and control arm of the trial. For less than 5% missingness overall, we intend to only carry out a complete-case analysis regardless of the mechanism behind the missingness. However, in order to gauge systemic differences in missingness, we intend to model missingness at endline as a function of baseline covariates, including treatment status. If there is a noticeable pattern in the missingness modelled in the previous step, multiple imputation may be required (in line with EEF guidance, 2022). If data is missing not at random, then multiple imputation is not appropriate as it is not able to generate unbiased estimates of the treatment effect. In this case, sensitivity analysis will be carried out and reported alongside the headline impact results.

Sensitivity analysis

The proposed baseline outcome in the secondary analysis is maths outcomes at KS2, which will be obtained through the NPD for all pupils included in the evaluation. The use of a baseline outcome measure is intended to account for variation in the dependent variable (i.e., the secondary outcomes), which would otherwise be left as heterogeneity in the residuals of the estimated model. However, we recognise that KS2 maths scores as a baseline outcome are not a direct historical measure of the secondary outcomes of self-efficacy and metacognition. Therefore, we will conduct a sensitivity analysis in which we will exclude the baseline KS2 scores from the secondary analysis, running an 'endline only' model. This will ensure that our results are robust to the inclusion of KS2 maths outcomes as a baseline in the secondary analysis, despite the dependent variable being self-efficacy or metacognition.

Implementation and process evaluation (IPE) design

Research questions

The IPE is designed to address the following research questions (and associated implementation dimensions) of CoachBright's Peer-to-Peer coaching programme, their relevance to the logic model is explained in the analysis section. Given the intervention is targeted at disadvantaged pupils, the IPE research questions do not outline specific barriers for this these pupils.

IPERQ1 - To what extent are School Coordinators and coaches able to access the training and support offered by the programme, including introductory onboarding and regular check-ins with CoachBright Programme Managers? (*responsiveness, implementation fidelity and moderators)*

IPERQ2 – To what extent have School Coordinators been able to select coaches and coachees and have they been appropriately matched? *(quality, reach)* Does the selection criteria support appropriate matching and are coaches motivated to deliver the Peer-to-Peer sessions? *(quality, responsiveness)*

IPERQ3 - Are peer to peer sessions delivered with fidelity and quality according to the TiDER model, and what are the key factors influencing deviations from the model? What are the barriers and facilitators to implementation? *(implementation fidelity, quality, adaptation)*

IPERQ4 - What does business as usual look like? Does this change in control schools during the intervention period? Are there any unintended consequences or changes in wider practices in schools (e.g. wider peer to peer tutoring in intervention settings)? *(monitoring of the control condition)*

IPERQ5 - Have coaches delivered one hour of coaching on a 1:1 basis over 10 weeks? *(Dosage)*

IPERQ6 - What are School Coordinators' perceived outcomes of the Peer-to-Peer coaching sessions on coaches and coachees and do these align with the coach and coaches' perceived outcomes? *(perceived impact)*

IPERQ7 - Has the programme facilitated coaches' aspirations, and what role do factors such as the university graduation ceremony and other programme elements play in supporting these aspirations?"

IPERQ8 - Are schools able to continue implementation with fidelity within the one year timeframe of the evaluation? *(implementation fidelity and responsiveness)*

IPERQ9 - What are the main barriers and facilitators to sustaining and scaling the programme within the one year timeframe of the evaluation? *(moderators)*

Following analysis of Cohort 1 data (relating to IPERQ 8 and 9) from the CoachBright-led model, an interim report will be produced (in June/July 2026) to inform delivery of the school-led model being piloted in Year 2.

Research methods

The embedded implementation and process evaluation in Year 1 of the study will take a mixed methods approach to data collection. This will include routine programme data, observations, in-depth interviews, focus groups and surveys. This approach to data collection will allow us to capture experiences and views of the School Coordinators and pupils involved in the trial.

We have mapped the research questions against key IPE dimensions below (Table 4). More details about each data collection method are provided below.

Routine data

The following routine programme data will be collected from CoachBright at the end of the programme for each cohort:

- Attendance at introductory onboarding for one School Coordinator and one member of the SLT per school collected during October 2025 (Cohort 1) and February 2026 (Cohort 2).
- Coach attendance at the two to three-hour face-to-face coach training event collected during October 2025 (Cohort 1) and February 2026 (Cohort 2).
- Coach and Coachee attendance at 50-to-60-minute weekly sessions for 10 weeks collected February 2026 (Cohort 1) and June 2026 (Cohort 2).
- Pupil Reflection Journals (selection from each school) collected February 2026 (Cohort 1) and June 2026 (Cohort 2).

Attendance data (a – c) will be used to understand responsiveness, dosage and programme implementation fidelity (IPERQ1, IPERQ5) and feasibility for the school-led model, as well as for the compliance analysis (compliance has been determined as at least 70% attendance at coach training and 70% of weekly coaching sessions (d). While feedback and impact report data (e) will allow a better understanding of implementation fidelity, including adaptations made to the programme, quality, and feasibility and adaptations needed for the school led model IPERQ2, IPERQ3 and IPERQ8).

Training

Two researchers from the evaluation team will attend one online onboarding session with school coordinators and SLTs and one launch event during October 2025 (Cohort 1). This will be done to support understanding of the process in which coaches and coachees are selected and to understand the main elements of the programme. *Two researchers will also attend 1 in person coach training session together to give us more understanding about what the coaches are*

expected to do. This will further support the evaluation team to develop appropriate data collection instruments.

Setting visits – observations and student focus groups

Setting visits will take place in schools (*n*=8) at midpoint for each cohort to observe two coaching sessions (*n*=16) and conduct two focus groups with four to six coaches (*n*=8) and coachees (*n*=8), respectively. The observation schedules will be designed to capture implementation of the core elements of the programme while the focus groups will capture coachers and coachees experiences of delivering/receiving the intervention. Four visits will take place during January/February 2026 (Cohort 1) and four visits will take place during May/June 2026 (Cohort 2). Two researchers will attend one setting together during Cohort 1 visits for quality assurance purposes. Setting visits will allow data to be gathered on coach motivation, implementation fidelity including barriers and facilitators at the individual level, as well as coach and coachee experiences (IPERQ2, IPERQ3, IPERQ5, IPERQ6, IPERQ7 and IPERQ9). Settings will be selected based on setting size and location. School Coordinators will be asked to select a range (ranging in motivation and prior maths attainment) of coaches and coachees for the focus groups. A Mentimeter[13] will be used to pose questions and facilitate engagement from the pupils.

Survey for coaches and coachees

Surveys will be distributed to Year 7 and Year 10 pupils in intervention (*n*=1,500) and control schools (*n*=1,500) at baseline following pupil selection (September/October 2025 Cohort 1 and February/March 2026 Cohort 2) and endline (February/March 2026 Cohort 1 and June/July 2026 Cohort 2) via Qualtrics. Pupils will use laptops or school computers to complete the survey. The survey will include measures of pupil attitudes towards maths, including some measures of maths enjoyment and maths anxiety taken from the Abbreviated Math Anxiety Scale (Hopko et al., 2003), given the relationship between maths anxiety, maths enjoyment and self-efficacy. The scale asks the respondents to rate how anxious they feel towards certain events e.g. thinking about an upcoming maths test one day before the test. Survey data will also be used to measure perceptions of training and the extent to which training supports coaches' understanding of the programme and their role within it (intervention schools only). At endline, the survey will gather data around support received, implementation fidelity and adaptations and aspirations and SSAT accreditation (intervention only) (IPERQ2, IPERQ3 IPERQ6, IPERQ7 and IPERQ9) as well as pupil attitudes towards maths and maths anxiety (control and intervention schools) (IPERQ4). Surveys will use a mix of closed and Likert scale questions and some open-ended questions.

School coordinator and SLT survey

[13] Mentimeter is an interactive software which allows you to pose questions. As the pupils answer this is displayed in a graph which can then be used to facilitate discussion.

An online survey will be developed for completion by school coordinators (*n*=100) and the designated member of the SLT (*n*=100) in the control and intervention groups at baseline (September/October 2025) and endline (February/March 2026 Cohort 1 and June/July 2026 Cohort 2), via Qualtrics. The baseline survey will be used to understand usual practice, including any programmes currently being implemented in schools and to understand schools' motivation for participation in the programme; while this is not explicitly in the ToC, evidence suggests that collective teacher 'buy-in' is a key factor in successful implementation in secondary schools (Dyssegaard et al., 2017). This data will also be used to identify schools to take part in the school-led model. At endline the survey will be designed to answer questions around other maths programmes/interventions being implemented in schools (IPERQ4) and will gather data for the cost evaluation. For intervention settings the survey will also include questions around training, implementation fidelity particularly around selecting coaches, perceived outcomes for coaches and coachees and the feasibility of the school-led model (IPERQ1, IPERQ2, IPERQ3, IPERQ6, IPERQ8 and IPERQ9). Surveys will use a mix of closed and Likert scale questions and some open-ended questions.

Interviews with intervention school coordinators

Interviews (*n*=12) with school coordinators (intervention schools only) will take place during January/February 2026 (Cohort 1) and May/June 2026 (Cohort 2) to gather data on training and support, selecting coaches, and the feasibility of the school led model (IPERQ1, IPERQ2, IPERQ3, IPERQ6, IPERQ8 and IPERQ9). Eight of the interviews will be conducted in settings who are selected for the observation visits and will be conducted face-to-face. A further four teachers will be selected for online interviews to give further breadth to the interview data (selected on setting size and location).

Focus groups with CoachBright Programme Managers

Focus groups with CoachBright programme managers (*n*=2, including 3 to 4 managers per focus group) will be conducted at Cohort 1 endline (January/February 2026) to assess training, barriers and facilitators to selecting coaches and coachees, coach motivation, and quality and fidelity of programme implementation (IPERQ1, IPERQ2, IPERQ3, IPERQ8 and IPERQ9). Programme managers will be selected randomly based on their availability.

Analysis

Table 4: IPE methods overview

IPE dimension	RQ addressed	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
Responsiveness	IPERQ1, IPERQ2	**Routine data**	Attendance and session quality data	50 schools (intervention)	Frequency counts Deductive coding

	, IPERQ8	**Surveys** (baseline/endline)	Online questionnaires	100 school coordinators, 100 SLT 1,500 Year 7 pupils, 1,500 Year 10 pupils	Descriptive statistics, t-tests, deductive coding, thematic analysis
		Interviews	Face to face/online interviews	12 school coordinators	Deductive coding, thematic analysis
		Observations /Focus groups	Face to face focus groups	8 schools 8 Year 7 groups, 8 Year 10 groups 2 programme manager groups	Count analysis Deductive coding, thematic analysis
Implementation Fidelity	IPERQ1, IPERQ3, IPERQ8	**Routine data**	Attendance and session quality data	50 schools (intervention)	Frequency counts Deductive coding
		Surveys (endline)	Online questionnaires (intervention)	50 school coordinators, 50 SLT 750 Year 7 pupils, 750 Year 10 pupils	Descriptive statistics, t-tests, deductive coding, thematic analysis
		Interviews	Face to face/online interviews	12 school coordinators	Deductive coding, thematic analysis
		Observations /Focus groups	Face to face focus groups	8 schools 8 Year 7 groups, 8 Year 10 groups 2 programme manager groups	Count analysis Deductive coding, thematic analysis
Moderators	IPERQ1, IPERQ9	**Routine data**	Attendance	50 schools (intervention)	Frequency counts Deductive coding
		Surveys (post)	Online questionnaires (intervention)	50 school coordinators, 50 SLT	Descriptive statistics, t-tests, deductive coding, thematic analysis

				750 Year 7 pupils, 750 Year 10 pupils	
		Interviews	Face to face/online interviews	12 school coordinators	Deductive coding, thematic analysis
		Observations /Focus groups	Face to face/ focus groups	8 schools 8 Year 7 groups, 8 Year 10 groups 2 programme manager groups	Count analysis Deductive coding, thematic analysis
Quality	IPERQ2 ,IPERQ 3,	**Routine data**	quality data	50 schools (intervention)	Frequency counts, Deductive coding
		Surveys (post)	Online questionnaires (intervention)	50 school coordinators, 50 SLT 750 Year 7 pupils, 750 Year 10 pupils	Descriptive statistics, t-tests, deductive coding, thematic analysis
		Interviews	Face to face/online interviews	12 school coordinators	Deductive coding, thematic analysis
		Observations /Focus groups	Face to face/online focus groups	8 schools 8 Year 7 groups, 8 Year 10 groups 2 programme manager groups	Count analysis Deductive coding, thematic analysis
Reach	IPERQ2	**Routine data**	quality data	50 schools (intervention)	Frequency counts Deductive coding
		Surveys (post)	Online questionnaires (intervention)	50 school coordinators, 50 SLT 750 Year 7 pupils, 750 Year 10 pupils	Descriptive statistics, t-tests, deductive coding, thematic analysis

		Interviews	Face to face/online interviews	12 school coordinators	Deductive coding, thematic analysis
		Observations /Focus groups	Face to face focus groups	8 schools 8 Year 7 groups, 8 Year 10 groups 2 programme manager groups	Count analysis Deductive coding, thematic analysis
Adaptation	IPERQ3	**Routine data**	quality data	50 schools (intervention)	Frequency counts Deductive coding
		Surveys (post)	Online questionnaires (intervention)	50 school coordinators, 50 SLT 750 Year 7 pupils, 750 Year 10 pupils	Descriptive statistics, t-tests, deductive coding, thematic analysis
		Interviews	Face to face/online interviews	12 school coordinators	Deductive coding, thematic analysis
		Observations /Focus groups	Face to face focus groups	8 schools 8 Year7 groups, 8 Year 10 groups 2 programme manager groups	Count analysis Deductive coding, thematic analysis
Monitoring of the control condition	IPERQ4	**Surveys** (pre/post)	Online questionnaires	100 school coordinators, 100 SLT	Descriptive statistics, t-tests, deductive coding, thematic analysis
Dosage	IPERQ5	**Routine data**	Attendance	50 schools (intervention)	Frequency counts
		Focus groups	Face to face focus groups	8 Yea 7 groups, 8 Year 10 groups	Deductive coding, thematic analysis
Perceived Impact	IPERQ6, IPERQ7	**Surveys** (post)	Online questionnaires (intervention)	50 school coordinators, 50 SLT	Descriptive statistics, t-tests,

				750 Year 7 pupils, 750 Year 10 pupils	deductive coding, thematic analysis
		Interviews	Face to face/online interviews	12 school coordinators	Deductive coding, thematic analysis
		Focus groups	Face to face focus groups	8 Year 7 groups, 8 Year 10 groups	Deductive coding, thematic analysis

IPERQ1 - To what extent are teachers and coaches able to access the training and support offered by the programme, including introductory meetings and regular check-ins with CoachBright trainers?

Measures and relation to Logic Model and causal mechanisms

To monitor access to training and support, endline survey data from school coordinators and members of the SLT from the intervention group will be triangulated with data from interviews with school coordinators. Attendance at the onboarding and attendance at coaching sessions will also be presented as a measure of teacher compliance. Coach endline survey data will also be triangulated with coach focus group data and attendance at the three-hour training event will be presented as a measure of coach compliance.

The school co-ordinator interview schedule will be designed to capture the teacher/school staff inputs of the logic model. It will specifically ask questions round the perceived usefulness of the onboarding meeting and if/how coordinators utilised the support from programme managers, including the regular check-ins. Similarly, coach focus group schedules will be designed to capture the coach inputs of the logic model and will ask questions around the perceived usefulness of the training and support from programme managers particularly in relation to resources and session planning.

The school-coordinator/SLT and coach endline survey will be developed around the same themes as the interview schedule (teacher/school staff/coach inputs of the logic model) and will also draw on any key themes taken from the interview data to ensure a fully triangulated approach.

Synthesis of data and analysis

Interview data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data around the teachers/school staff inputs of the logic model. Survey data will be analysed descriptively. Interview data will be triangulated with data collected from the survey. The survey data and interview data around

training will also be presented with attendance at the onboarding sessions/three-hour training and coaching sessions to give a full picture the responsiveness of the school coordinators/SLT and coaches towards training and to see if this level of responsiveness acts as a moderator to support the outputs of the logic model (IPERQ2).

IPERQ2 – To what extent have teachers been able to select coaches and coachees and have they been appropriately matched? Does the selection criteria support appropriate matching and are coaches motivated to deliver the Peer-to-Peer sessions?

Measures and relation to Logic Model and causal mechanisms

To assess the extent to which school coordinators are able to select coaches and coachees using the training and guidance provided, data from all endline surveys (school co-ordinator, SLT, coaches and coachees) will be triangulated with school-coordinator interview data and pupil and programme manager focus group data. Measures will be designed to specifically ask about the ease/difficulty of pupil selection.

To address whether the guidance criteria support appropriate pupil matching, school-coordinator endline surveys and interviews will be developed with targeted questions and the data from each of the sources will be analysed simultaneously.

To assess whether coaches are motivated to deliver the sessions, data from the endline coach survey and coach focus group measures will ask direct and indirect questions around motivation. Data from both sources will be used to gain a deeper perspective of coach motivation (given that those attending focus groups will probably be more motivated than those who do not attend).

Synthesis of data and analysis

Interview data and focus group data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data around the teachers/school staff and coach outputs of the logic model. Survey data will be analysed descriptively.

Appropriateness of coach and coachee selection will combine the data from each of the data collection tools and sources (coaches and coachees, school coordinators, SLT and Programme Manager) to understand the different views of all those involved in implementation both within groups (e.g. coaches) and across groups (e.g. coaches and school-coordinators views). As the appropriateness of matching may be moderated by whether or not the quality of the guidance provided supports appropriate matching, the data will be presented together. Additionally, if the IPERQ1 data shows differences between schools in access to training and coaching further analysis will be conducted to see if attendance at training and the amount of support accessed, impacts on appropriate matching of coaches to coachees.

Coach motivation data from the endline survey will be triangulated with coach focus group data to understand responsiveness of the coaches. If access to training and support is found to influence the appropriateness of matching it is possible that this may have a knock-on effect on coach motivation. In this case, further analysis will be carried out to assess the potential influence of training on motivation i.e. analysis will be performed separately on schools where matching is in line with the logic model and schools where matching is not in line with the logic model. If differences in coach motivation are found between schools (i.e. at the school-level), this will then inform further analysis as coach motivation is likely to influence the quality of implementation fidelity (IPERQ3), compliance (IPERQ5), perceived outcomes (IPERQ6) and coach aspirations (IPERQ7).

IPERQ3 - Are peer to peer sessions delivered with fidelity and quality according to the TiDER model?

Measures and relation to Logic Model and causal mechanisms

To understand how the programme impacts on the short and longer-term outcomes it is important to know if the programme has been implemented with high fidelity and quality. To understand what high quality and fidelity should look like researchers will attend the training sessions held for the coaches. A quality and fidelity framework will then be produced by the evaluation team and shared with the delivery team for feedback.

The quality and fidelity framework will be used to produce an observation schedule which focuses directly on these two areas. If any obvious areas of low quality or fidelity are seen in the observations, or if adaptations appear to have been made, these will be incorporated into the school co-ordinator interviews and coach/coachee focus groups. The endline surveys (school co-ordinator/SLT and coach and coachee) will also ask questions specifically around implementation of the core programme elements along with questions around barriers and facilitators of implementation.

Synthesis of data and analysis.

Interview data, focus group data and qualitative observation data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data around implementation. Survey data will be analysed descriptively and will be used to corroborate (or oppose) qualitative data.

If coach motivation (IPERQ2) is found to differ between schools, analysis at the school-level will be used to identify whether or not coach motivation influences implementation fidelity and quality.This additional analysis will look at varying contextual factors across schools.

IPERQ4 - What does business as usual look like? Does this change in control schools during the intervention period? Are there any unintended consequences or changes in wider practices in schools (e.g. wider peer to peer tutoring in intervention settings)?

Measures and relation to Logic Model and causal mechanisms

To understand business as usual, school-coordinator baseline and endline surveys will be analysed to understand if any similar approaches are being implemented in control settings (monitoring of the control condition) and to assess any changes from baseline to endline in control and intervention settings. Surveys will be developed to also monitor implementation of similar programmes (Peer-to-Peer coaching) or implementation of programmes targeted at improving maths attainment for pupils in either year 7 or year 10, in control schools. This is an essential element of the IPE as it seeks to understand any influence similar programmes may have on pupil outcomes. Additionally, the surveys will ask about unintended consequences or changes in wider practice. In intervention schools unintended consequences may be the implementation of the Peer-to-Peer coaching model in other years or in other subjects or not implementing a programme or intervention in favour of Peer-to-Peer coaching.

Synthesis of data and analysis.

Survey data will be analysed descriptively, and comparative analysis will be conducted between control and intervention schools and baseline and endline (control and intervention settings).

IPERQ5 - Have coaches delivered one hour of coaching on a 1:1 basis over 10 weeks?

Measures and relation to Logic Model and causal mechanisms

Routine data on the number of one-to-one sessions delivered over the 10-week period will be collected from CoachBright programme officers, along with records of the length of sessions completed. The dosage will be used to assess compliance with at least 70% of one-hour sessions attended for a minimum duration of 45 minutes being classed as compliant. The data will be used to understand the potential influence of high/low attendance at coaching sessions on pupil outcomes.

Attendance data will be triangulated with coach/coachee focus group data which will be designed to ask specifically about barriers and influences to attending the one-to-one sessions. In addition, the focus groups with be designed to gather data around reflection of academic progress and if the sessions have allowed coachees to practice specific topics i.e. topics which match the learning needs of the coachees, or not.

Synthesis of data and analysis.

Compliance data on the number of sessions attended will be converted into percentages at the school-level. Compliance data on the length of sessions will be analysed descriptively at the school-level. If coach motivation (IPERQ2) is found to differ at the school-level appropriate additional analysis will be conducted between low-level and high-level coach motivation schools. Coach and coachee focus group data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data around dosage. The qualitative data will be presented alongside compliance data.

IPERQ6 - What are teachers perceived outcomes of the Peer-to-Peer coaching sessions on coaches and coachees and do these align with the coach and coachees perceived outcomes?

Measures and relation to Logic Model and causal mechanisms

Perceived impacts on year 7 and year 10 pupils are divided into short-term and long-term outcomes on the logic model. Data will be collected from endline school co-ordinator/SLT surveys and school co-ordinator interviews and coach and coachee baseline/endline surveys and focus groups. The tools will be designed to capture, for coaches, self-reported improvements in the ability to support year 7 pupils maths attainment and improvements in leadership skills. For coachees, the tools will be designed to capture potential increases in enjoyment and interest in maths, improved confidence in own maths ability, and improved school attendance.

For both coaches and coachees improvement in metacognitive skills and self-efficacy will be measured through the impact evaluation. However, the IPE will also measure engagement in maths lessons along with maths anxiety through the Abbreviated Math Anxiety Scale which will be embedded in baseline and endline surveys.

Synthesis of data and analysis.

Interview data, focus group data and qualitative survey data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data around perceived impact with a specific focus on the areas outlined above. Survey data will be analysed descriptively and will be used to corroborate (or oppose) qualitative data.

Data taken from the abbreviated maths scale will be analysed using t-tests to measure differences between control and intervention group pupils and differences at baseline and endline between and within control and intervention group pupils.

If coach motivation (IPERQ2) is found to differ between schools, analysis at the school-level will be used to identify whether or not coach motivation influences implementation fidelity and quality.

IPERQ7 - Has the programme facilitated coaches aspirations and is this supported by the university graduation ceremony?

Measures and relation to Logic Model and causal mechanisms

Perceived impact on coaches in terms of their aspirations will be collected from baseline and endline coach surveys and coach focus groups. Specifically, the surveys and focus groups schedules will be designed to ask questions about their future aspirations and attending

university graduation (baseline and endline surveys and focus groups) and will also ask about the type of SSAT accreditation (bronze, silver, gold) they will/have applied for.

Synthesis of data and analysis.

Focus group data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data around perceived impact with a specific focus on the areas outlined above. Survey data will be analysed descriptively and will be used to validate (or dispute) qualitative data. Data around coach aspirations will be analysed to look for change from baseline to endline.

If coach motivation (IPERQ2) is found to differ between schools, analysis at the school-level will be used to identify whether or not coach motivation influences implementation fidelity and quality.

The following two research questions will be analysed as part of the main evaluation report and will also be used to produce the interim report to guide delivery of the school-led model. As such, the two separate areas of focus (CoachBright-led model and School-led model) are outlined clearly below:

IPERQ8 - Are schools able to continue implementation with fidelity?

Measures and relation to Logic Model and causal mechanisms

As part of the CoachBright led model this research question is designed to understand whether schools feel they are able to continue implementing the Peer-to-Peer coaching model within their school. This relates to the short-term and longer-term teacher/school staff outcomes in the logic model and in general, the responsiveness to the programme. It will also ask questions around the cost of the CoachBright-led model and costs around the school-led model. This data will be collected through school-coordinator/SLT endline surveys and school-coordinator focus groups.

This research question is designed to understand if schools feel they would be able to implement the school-led model with high fidelity without the need for facilitation from the CoachBright programme manager. It is closely linked with IPERQ9. This data will be collected through endline school-coordinator/SLT endline surveys and school-coordinator focus groups. The survey and interview schedule will be designed to ask questions around continued implementation of the core elements of the programme. They will also ask school-coordinators/SLTs to reflect on whether they feel implementation of the CoachBright led model is a prerequisite to enable implementation of the school-led model.

Synthesis of data and analysis.

Interview data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data around continued implementation. Survey data will be analysed descriptively and will be presented alongside interview data.

For reflection of the school-led model the data will be analysed and presented in such a way to be informative for the pilot trial (Year 2). It will present which core elements can be continued to be implemented with high fidelity and any which may be more difficult to continue implementing with high fidelity.

IPERQ9 - What are the main barriers and facilitators to sustaining and scaling the programme? (*moderators*)

Measures and relation to Logic Model and causal mechanisms

As part of the CoachBright led model this research question is designed to understand the main barriers to, and facilitators of sustaining the Peer-to-Peer coaching model for maths in schools. It will have a particular focus on the role of the senior leadership team and if/how this moderates facilitation. Data will be collected through the programme manager focus groups, school co-ordinator interviews and school-coordinator/SLT endline surveys. Main barriers and facilitators discussed during the interviews will be included in the endline surveys to gather more information.
This research question is designed to understand perceived barriers to, and facilitators of facilitating the school-led model. It is closely linked with IPERQ8 in that it will seek to understand implementation with fidelity and how this can be achieved. Data will be collected through the programme manager focus groups, school co-ordinator interviews and school-coordinator/SLT endline surveys. The instruments will be designed to ask reflective questions the school-led model.

Synthesis of data and analysis

Interview data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data around barriers and facilitators of sustaining implementation/school-led model implementation. Survey data will be analysed descriptively and will be presented alongside interview data.

Reflecting on the school-led model, the data will be analysed and presented in such a way to be informative for the pilot trial (Year 2). It will present key perceived barriers and facilitators in terms of programme delivery.

Interim report

The interim report will be presented to the delivery team and EEF as a presentation (June/July 2026). It will present all qualitative and quantitative data around perceived implementation fidelity of the school-led model and perceived barriers and facilitators (as outlined above). The findings will be presented alongside recommendations for both implementation of the school-led model and for the pilot study design.

Cost evaluation design

We will evaluate the costs of implementing the P2P intervention using data gathered through interviews and surveys administered to school coordinators across all schools. In schools assigned to receive the treatment, we will evaluate the cost of implementation. In schools assigned to the control we will evaluate the costs of business as usual, including both the direct and indirect costs of running programmes which are similar to P2P.

We will follow EEF guidelines for calculating costs (EEF 2023). The cost of implementation will be calculated per child over the course of the intervention period. The aspects of costs incurred by settings that will be gathered through our data collection tools include: direct costs of running the programme, practitioner time used for training (as staff cover may be required to allow School Coordinators to undertake onboarding), preparation and delivery of the programme, supplemental material cost incurred to deliver the programme, additional staff time used to support delivery of the programme (e.g. the need to employ other cover staff to allow practitioners time to prepare the activities). Costs such as time spent, stationery and other supplies will be monetised using market estimates. We will use sensitivity analysis to account for heterogeneity of costs between settings. The evaluation will measure pre-requisite costs, start-up costs (e.g. training), and recurring costs (e.g. costs of materials, staff time required for support).

Ethics and registration

The trial will be registered on the International Standard Randomised Controlled Trial Number (ISRCTN) registry, which is used to describe randomised controlled trials (RCTs) and efficacy trials at inception. Once registered, this protocol will be updated with the assigned registration number.

The ethics and registration processes are in accordance with the ethics policies adopted by RAND Europe and the University of Leeds. The study received ethical approval from RAND Europe's internal review board on 24/02/2025 (ref: 022807.022). It has also received ethical approval from the University of Leeds Cross-Faculty Research Ethics Committee (Faculties of Business, Environment & Social Sciences) (Ref: 2621)

Prior to pupils' data being sent to the delivery team, parents will be sent information sheets and withdrawal forms by the setting and will have the opportunity to return these. Parents can withdraw their children at any time from the data collection activities. If parents choose to withdraw their children from data collection, their data will not be collected or will be deleted, as appropriate.

RAND Europe will collect consent forms for all pupils, school coordinators, practitioners, and programme managers that participate in an interview. The front page for each online survey will contain a privacy notice informing respondents that participation in the survey is entirely voluntary. The consent form in the survey will be built into the data collection tool so that those moving past a certain page (following the privacy notice and information on the research) will have given consent for the data to be used in the research.

None of the evaluation team has any conflicts of interest and all members of the study team have approved this protocol prior to publication.

Data protection

The RAND Europe and University of Leeds evaluation team have extensive experience handling personal data. Many of our researchers are accredited by the Office for National Statistics to use data from the National Pupil Database.

For the duration of the evaluation, RAND Europe, the University of Leeds, and CoachBright will act as joint data controllers and will be the main point of contact for any matters relating to the protection of all personal data once the recruitment is completed and will make decisions about how and what personal data is used in the evaluation. This means all teams are responsible for deciding the purpose and legal basis for processing data.

The legal basis for RAND and CoachBright is "legitimate interest", as a legitimate interest assessment (LIA) has been conducted.[14] Legitimate interest is an appropriate basis because the data collected as part of this evaluation will be used in ways that people would reasonably expect (i.e. for the benefit of improving support for executive functioning and early mathematical development in children) and that have minimal privacy impact. Legitimate interests apply where processing is necessary for the purpose of the legitimate interest pursued by the controller (see GDPR Article 6 (1) (f)) and for statistical and research purposes (See GDPR Article 89). CoachBright (the delivery team) rely on public interest as the legal basis for use of the data.

The legal basis for the University of Leeds is where it is necessary for the performance of a task carried out in the public interest as set out in Article 6(1)(e) of the UK GDPR. The specific legislation which allows this is Section 10 of the Education Act 1996. The legal basis for processing special category data is for reasons of substantial public interest as detailed in Article 9(2)(g) of the UK GDPR.

This study will be conducted in compliance with current data protection legislation. As such, all child data and any other personal data used for the project will be treated with the strictest confidence and will be used and stored in accordance with the General Data Protection Regulation (GDPR) (2018) and the Data Protection Act (2018).

The evaluation team will use basic identifiers (e.g., name and date of birth) to associate information with individuals to create datasets for research purposes. The rights and freedoms of the subjects will not be affected as information will only be identifiable during processing to the evaluation and delivery teams and not otherwise. If parents choose to withdraw their children from data collection, the data will not be collected, or will be deleted if already collected.

[14] For more information about legitimate interest, please see: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/.

The evaluation team have put appropriate security measures in place to keep personal data secure and to prevent any unauthorised access to or use of it. The evaluation team will collect and store all evaluation data in accordance with the Data Protection Act (2018) and GDPR requirements. Evaluation data will be stored on secure servers. Data transferred between the delivery and evaluation teams will be encrypted or use secure file transfer protocols.

Personnel

Delivery team: Coach Bright

Project manager and Director of Programmes: Beth Gascoigne-Owens

CEO: Joe McGinn

DPO: George Shackleton

Data and Impact Manager: Maddie Kilgarriff

Team leader for the Programme Managers: Jade Hargreaves

Evaluation team: RAND Europe and University of Leeds

Principal investigator and project leader: Elena Rosa Speciani

Project manager: Martha Aitken

Impact evaluation lead: James Merewood

Implementation and process evaluation and data collection: Erin Dysart and Louise Tracey

Risks

Risks	Assessment (Likelihood/Impact)	Mitigation strategy	Impact post-mitigation
Recruitment failure	Likelihood: low Impact: high	This can be mitigated by regular dialogue over any recruitment issues and ensuring that the design introduces minimal burden to schools. Timelines have been agreed to ensure there is adequate time for all activities.	Moderate
Attrition	Likelihood: moderate Impact: high	We propose recruiting more schools (in a waitlist) to build in a buffer for attrition of child and school level. Schools will be given clear information about participation before signing up. Test burden will be kept low to maximise participation and reduce drop out.	Moderate

		Waitlist settings will be contacted in the event of a drop out, so that schools can be replaced.	
Small number of FEEE children for analysis	Likelihood: moderate Impact: moderate	We propose that the delivery team aim to recruit schools in areas of high deprivation to support analysis of FSM children in both the impact and IP evaluations	Low
Low participation rates for IPE surveys and interviews	Likelihood: moderate Impact: moderate	Sufficient data collection window given with real-time monitoring of response rates to allow for reminders to be targeted.	Low
Quality of reporting	Likelihood: moderate Impact: moderate	Applying RAND QA processes, including expert review. PI with considerable experience of EEF reporting standards	Low
Lack of coordination between Evaluation Team, the EEF and the Delivery team	Likelihood: moderate Impact: moderate	All teams have attended initial meetings and have agreed on roles and responsibilities. Regular contact between key team members from each organisation will be maintained throughout	Low
Evaluation team members' absence or turnover	Likelihood: moderate Impact: low	The team can be supplemented by researchers with considerable experience in evaluation from the larger RAND Europe pool.	Low

Timeline

Table 6: Timeline

Dates	Activity	Staff responsible/ leading
March – July 2025	Recruitment of settings	Delivery team
April – July 2025	MoUs received for settings involved	Delivery team
May – Aug' 2025	Pilot Measures	RAND Europe
Sept' – Oct' 2025	Baseline data collection	University of Leeds
Sept' – Oct' 2025	Baseline survey for teachers	University of Leeds

Oct' 2025	Randomisation	RAND Europe
Nov' 2025 – Feb' 2026	Training and delivery of programme in settings (cohort 1)	Delivery team
Jan' – Feb' 2026	Endline data collection (cohort 1)	University of Leeds
Feb' – June 2026	Training and delivery of programme in settings (cohort 2)	Delivery team
May – June 2026	Endline data collection (cohort 2)	University of Leeds
Aug' – Nov' 2026	IPE and IE analysis	University of Leeds and RAND Europe
Nov' 2026	IPE and IE report writing	University of Leeds and RAND Europe
May 2026	Recruitment begins for the school-led model	CoachBright

References

Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin. G. (2019a). Peer tutoring in mathematics in primary education: a systematic review. *Educational Review*, 71(6), 767-791. https://doi.org/10.1080/00131911.2018.1474176

Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin. G. (2019b). Peer tutoring and mathematics in secondary education: literature review, effect sizes, moderators, and implications for practice. *Heliyon*. 5. https://doi.org/10.1016/j.heliyon.2019.e02491.

Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin. G. (2025). A Meta-analysis of the effect of peer tutoring in Science, Technology, Engineering and Mathematics (STEM) subjects. *International Journal of Educational Research Open*, 8, 100446. https://doi.org/10.1016/j.ijedro.2025.100446

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. *Sage Publications*, Inc

Cropp, I. (2017). Using peer mentoring to reduce mathematical anxiety. *Research Papers in Education*, 32(4), 481–500. https://doi.org/10.1080/02671522.2017.1318808

Culora, A., Dimova, S., Ilie, S., Sutherland, A., & Gilde, L. (2020). *Peer Assisted Learning Strategies – UK: Evaluation Report*. Education Endowment Foundation.

De Backer, L., Van Keer, H. & Valcke, M. (2016). Eliciting Reciprocal Peer-Tutoring Groups' Metacognitive Regulation Through Structuring and Problematizing Scaffolds. *The Journal of Experimental Education*. https://doi.org/10.1080/00220973.2015.1134419

Department for Education. (2021). *National Tutoring Programme: Guidance for schools*. GOV.UK. https://www.gov.uk/government/publications/national-tutoring-programme-guidance-for-schools

Department for Education. (2023). *Maths to 18: The Prime Minister's plan to transform maths education*. GOV.UK. https://www.gov.uk/government/news/maths-to-18-the-prime-ministers-plan-to-transform-maths-education

Department for Education. (2024). *Independent review of tutoring in schools: Phase 1 findings*. GOV.UK. https://www.gov.uk/government/publications/independent-review-of-tutoring-in-schools-and-16-to-19-providers/independent-review-of-tutoring-in-schools-phase-1-findings

Demack, S., McCaig, C., Wolstenholme, C., Stevens, A., & Fumagalli, L. (2016). *ThinkForward: Evaluation report and executive summary.* Education Endowment Foundation.

Demack, S., Culliney, M., Boyan, M., Wolstenholme, C. (2022). Realistic Maths Education: Evaluation Report. Education Endowment Foundation. Link.

Department of Education (2021). *Education Recovery Plan*. Retrieved from: https://www.gov.uk

Department of Education (2023). *Pupil Premium: Conditions of Grant*. Retrieved from: https://www.gov.uk

Donker, A.S., Boer, der H, Kostons, D. Dignath van Ewijk, C.C., & van der Werf, M.P.C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis, Educational Research Review, 11, 1 – 26. https://doi.org/10.1016/j.edurev.2013.11.002.

Dyssegaard, C.B., Egelund, N., Sommersel, H.B (2017) A systematic review of what enables or hinders the use of research-based knowledge in primary and lower secondary school. Copenhagen: Danish Clearinghouse for Educational Research, Department of Education, Aarhus University.

EEF (2021). *Teaching and Learning Toolkit: Peer Tutoring*. Retrieved from: Education Endowment Foundation

EEF (2022). Statistical analysis guidance for EEF evaluations. Retrieved from: Education Endowment Foundation.

EEF (2023). Cost Evaluation Guidance for EEF Evaluations. Retrieved from: Education Endowment Foundation.

Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology*, 84(3), 331–339. https://doi.org/10.1037/0022-0663.84.3.331

Fatma, A. C. A. R., & Engin. A. D. E. R. (2023). Metacognition Used by Tutors during Peer Tutoring Sessions in Mathematics. *Elementary Education Online*, *16*(3), 1185–1200. Retrieved from https://ilkogretim-online.org/index.php/pub/article/view/385

Fisher, D. (2001). Cross Age Tutoring: Alternatives to the Reading Resource Room for Struggling Adolescent Readers. *Journal of Instructional Psychology*, 28(4), 234. https://www.proquest.com/scholarly-journals/cross-age-tutoring-alternatives-reading-resource/docview/1416362834/se-2

Gaustad, J. (1993). *Peer and cross-age tutoring*. ERIC Digest. University of Oregon. https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/3300/digest079.pdf

Gardner, W. E. (1973). The Effects Of Intergrade Tutoring With Group Guidance Activities On The Reading Achievement, Self-concept, Attitudes Toward School And Behavior Of Third-grade And Fourth-grade Tutors And On The Reading Achievement And Behavior Of First-grade And Second-grade Tutees (Order No. 7331723).

Ginsburg-Block, M., & Fantuzzo, J. (1997). Reciprocal peer tutoring: An analysis of "teacher" and "student" interactions as a function of training and experience. *School Psychology Quarterly*, 12(2), 134–149. https://doi.org/10.1037/h0088955

Hopko, D.R., Mahadevan, R., Bare, R.L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, Validity, and Reliability. *Assessment*, 10, 178-182.

Husain, F., Bartasevicius, V., Marshall, L., Chidley, S. & Forsyth, E. (2019). Fit to Study: Evaluation Report. Education Endowment Foundation. Link.

ImpactEd (2023). CoachBright: Summary Impact Report. Link.

Jerrim, J., Austerberry, H., Crisan, C., Ingold, A., Morgan, C., Pratt, D., Smith, C. & Wiggins, M. (2015). Mathematics Mastery: Secondary Evaluation Report. Education Endowment Foundation. Link.

Liu, L., Saeed, M. A., Abdelrasheed, N. S. G., Shakibaei, G., & Khafaga, A. F. (2022). Perspectives of EFL learners and teachers on self-efficacy and academic achievement: The role of gender, culture and learning environment. *Frontiers in Psychology, 13*, Article 996736. https://doi.org/10.3389/fpsyg.2022.996736

Lloyd, C., Edovald, T., Kiss, Z., Morris, S., Skipp, A., & Ahmed, H. (2014). Paired Reading: Evaluation Report and Executive summary. Education Endowment Foundation.

Lynch, S., Bradley, E., Aston, K., Schwendel, G., & Lord, P. (2024). *National Tutoring Programme – Year 4: Evaluation and reflections*. National Foundation for Educational Research.

McKinstery, J. & Topping, K. J. (2003). Cross–age Peer Tutoring of Thinking Skills in the High School. *Educational Psychology in Practice*, *19*(3), 199–217. https://doi.org/10.1080/0266736032000109465Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. Journal of Educational Psychology, 86(2), 193-203

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2013). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology, 34(1), 29–48. https://doi.org/10.1080/01443410.2013.797339*

Pro Bono Economics. (2021). *Counting on the recovery: the role for numeracy skills in 'levelling up' the UK*. https://www.nationalnumeracy.org.uk/sites/default/files/documents/Counting-on-the-Recovery.pdf

Raabe, I. J., Boda, Z., & Stadtfeld, C. (2019). The Social Pipeline: How Friend Influence and Peer Exposure Widen the STEM Gender Gap. *Sociology of Education*, 92(2), 105-123. https://doi.org/10.1177/0038040718824095 (Original work published 2019)

Ray, D., Muñoz, A., Zhang, M., Li, X., Chatterjee, N., Jacobson, L. P., & Lau, B. 2022. Meta-analysis under imbalance in measurement of confounders in cohort studies using only summary-level data. *BMC Medical Research Methodology*, 22, 143. https://doi.org/10.1186/s12874-022-01614-9

Riegle-Crumb, C., Farkas, G., & Muller, C. (2006). The Role of Gender and Friendship in Advanced Course Taking. *Sociology of Education*, 79(3), 206-228. https://doi.org/10.1177/003804070607900302 (Original work published 2006)

Roberts, N., & Spangenberg, E. D. (2020). Peer tutors' perceptions of their role in motivating learners to engage with mathematics. *Pythagoras, 41*(1), a528. https://doi.org/10.4102/pythagoras.v41i1.528

Robinson, D.R., Schofield, J.W. & Steers-Wentzell1, K.L. Peer and Cross-Age Tutoring in Math: Outcomes and Their Design Implications. *Education Psychology Review* **17**, 327–362 (2005). https://doi.org/10.1007/s10648-005-8137-2

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4), 534–574. https://doi.org/10.3102/0034654307309920Sharp, R. (2021). Action research: Coaching to build maths students' confidence. *The Education and Training Foundation*. Link.

Singh, A., Uwimpuhwe, G., Vallis, D., Akhter, N., Coolen-Maturi, T., Higgins, S., Einbeck, J., Culliney, M., & Demack. S. (2023). Improving Power Calculation in Education Trials.

Social Mobility Commission (2024) Attainment at age 16. Retrieved from: https://www.gov.uk

Sperling, R. A., Howard B. C., Miller L. A., & Murphy C. 2002. Measures of Children's Knowledge and Regulation of Cognition. *Contemporary Educational Psychology*, 27(1), pp. 51-79. https://doi.org/10.1006/ceps.2001.1091

Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2012). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation*, 8(2), 117-139.

Smith, E., & Andersen, I. G. (2022). Do Same-Gender Peers in the Classroom Have Heterogeneous Impacts on Male and Female Students? *Socius*, 8. https://doi.org/10.1177/23780231221105378 (Original work published 2022.)

Tenenbaum, H. R., Winstone, N. E., Leman, P. J., & Avery, R. E. (2019). How effective is peer interaction in facilitating learning? A meta-analysis. *Journal of Educational Psychology, 112*(7), 1303–1319. https://doi.org/10.1037/edu0000436

Topping, K. J. (2005). Trends in Peer Learning. *Educational Psychology*, 25(6), 631–645. https://doi.org/10.1080/01443410500345172

Topping, K., Miller, D., Thurston, A., Mcgavock, K., & Conlin, N. (2011). Peer tutoring in reading Scotland: Thinking big. *Literacy*, 45(1), 3-9. https://doi.org/10.1111/j.1741-4369.2011.00577.x

Topping, K. J., Peter, C., Stephen, P., & Whale, M. (2004). Cross-age peer tutoring of science in the primary school: influence on scientific language and thinking. Educational Psychology, 24(1), 57–75. https://doi.org/10.1080/0144341032000146449

Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K., & Miller, D. (2011). Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement*, 22(3), 265-289. https://doi.org/10.1080/09243453.2011.589859

Usher, E. L., & Pajares, F. (2009). *Sources of Middle School Mathematics Self-Efficacy Scale* [Database record]. APA PsycTests. https://doi.org/10.1037/t21546-000

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. (2003). Why do only some events cause learning during human tutoring?. *Cognition and Instruction*, 25(4), 1–63. https://www.jstor.org/stable/3233810

Wang, X., Houang, R. T., Schmidt, W., H., & Kelly., K. S. 2024. Relationship Between Opportunity to Learn, Mathematics Self-Efficacy, and Math Performance: Evidence from PISA 2012 in 63 Countries and Economies. *International Journal of Science and Mathematics Education, 22, pp. 1683-1708*. *https://doi.org/10.1007/s10763-024-10446-6*

Webb, N. M., Mastergeorge, A. 2003. Promoting effective helping behaviour in peer-directed groups. *International Journal of Education Research*, 39(1-2), pp. 73-97. https://doi.org/10.1016/S0883-0355(03)00074-0

White, A. H. (2019). The Impact of Implementing an Individualised Peer-Tutoring Programme on Low-Attaining Secondary School Readers. https://doi.org/10.17863/CAM.52044

Yang, E. F. Y., Chang, B., Cheng, H. N. H. & Chan, T. (2016). Improving Pupils' Mathematical Communication Abilities Through Computer-Supported Reciprocal Peer Tutoring. *Educational Technology & Society,* 19(3), 157-169. http://www.jstor.org/stable/jeductechsoci.19.3.157

Zimmerman, B. J. (2000). *Self-efficacy: An essential motive to learn*. Contemporary Educational Psychology, 25(1), 82-91