Education Endowment Foundation

Independent evaluation of the Oxford Teaching Effective Early Mathematics and Understanding in Primary Schools (TEEMUP) professional development programme: A two-armed cluster randomised controlled trial

Evaluation report

December 2025

Lyn Robinson-Smith, Caroline Fairhurst, Katie Whiteside, Imogen Fountain, Jess Hugill-Jones, Tom Davill, Kerry Bell, Carole Torgerson and David

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

We do this by:

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

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

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

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

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Department for Education

What Works Network

Table of contents

About the evaluator

The project was independently evaluated by a team from the University of York:

Lyn Robinson-Smith; Caroline Fairhurst; Katie Whiteside; Imogen Fountain; Jess Hugill-Jones; Tom Davill; Kerry Bell; Hannah Ainsworth; Louise Elliott; Heather Leggett; Kalpita Baird; Rachel Bottomley-Wise; Emma Standley; Carole Torgerson; and David Torgerson. The co-principal investigators were Dr Lyn Robinson-Smith and Hannah Ainsworth until August 2022, and Dr Lyn Robinson-Smith and Caroline Fairhurst from August 2022.

Contact details:

Dr Lyn Robinson-Smith,
York Trials Unit,
Department of Health Sciences,
University of York,
Heslington, York,
YO10 5DD

Email: lyn.robinson-smith@york.ac.uk

Acknowledgements

The evaluation team would like to thank all the schools, teachers, and pupils for giving their time and efforts to participate in the evaluation of TEEMUP (Teaching Effective Early Mathematics and Understanding in Primary Schools).

We would also like to thank the delivery team from University of Oxford for their co-operation in all aspects of this evaluation.

The evaluation team would like to acknowledge the patience of all involved with this trial, including the Education Endowment Foundation, in relation to delays and disruption caused by the COVID-19 pandemic.

This work was undertaken in the Office for National Statistics Secure Research Service using data from ONS and other owners and does not imply the endorsement of the ONS or other data owners.

Executive summary

The project

Teaching Effective Early Mathematics and Understanding in Primary Schools (TEEMUP) is a maths and self-regulation professional development (PD) training programme for Reception and Year 1 teachers. TEEMUP aims to develop educators' content knowledge, pedagogical knowledge, and pedagogical content knowledge with examples of high-quality practice and classroom resources to support the teaching of maths. The programme also improves how educators support the development of self-regulation and behaviour for learning of children, as well as a focus on improving teacher confidence, self-assessment, and self-reflection. It is designed to equip teachers with the tools to tailor their teaching to their school's context and resources, rather than providing a scripted set of activities. TEEMUP was developed and delivered by a team based at the University of Oxford. Participation in TEEMUP requires teachers to attend training across two full days and seven afternoon workshops, a minimum of three specialist coaching sessions, and access to a dedicated website with PD resources and additional materials.

This evaluation was a two-armed, cluster randomised controlled trial (RCT). Schools in the trial were randomly allocated to have up to three teachers participate in the PD or continue with teaching as usual. A total of 93 schools and 2,567 pupils from two Reception cohorts participated in the evaluation. A complementary implementation and process evaluation (IPE) explored how the programme was implemented across different schools, as well as barriers and facilitators to implementation. This involved staff surveys, interviews with school staff and the delivery team, observations of training workshops, and monitoring data.

The trial was due to start in 2020, but the start was disrupted by the COVID-19 pandemic. This meant delivery ran from January 2022 to May 2023, with Cohort 1 beginning Reception in September 2021 and Cohort 2 beginning Reception in September 2022.

Table 1: Key conclusions

Key conclusions	
1.	Pupils in TEEMUP schools made the equivalent of one additional month's progress in maths attainment at the end of Year 1, on average, compared to pupils in other schools. These results have a high-security rating.
2.	Free school meals (FSM)-eligible pupils in TEEMUP schools made the equivalent of two additional months' progress in maths attainment compared to pupils eligible for FSM in other schools. These results may have a lower security than the overall findings because of the smaller number of pupils.
3.	Teachers observed marked improvements in self-regulation and personal, social, and emotional development (PSED) outcomes for pupils, although it is unclear whether these were sustained in the longer term.
4.	Findings indicate that the TEEMUP PD was well received by teachers, who perceived a change in their own teaching and confidence. Teachers also perceived improvements to pupils' confidence in maths as well as a positive impact on children's self-regulation and behaviour within the classroom.
5.	The delivery of the programme was impacted by the COVID-19 pandemic and key elements of the training were adapted as a result. Competing commitments, which were exacerbated by the pandemic limited the extent to which teachers could attend workshops and follow requirements of the programme.

EEF security rating

These findings have a high-security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools. The trial was a well-designed two-armed, cluster RCT. Around 15.1% of the children who started the trial were not included in the final analysis, mostly because children had left school by the end of the programme or were absent on the day of testing.

Additional findings

Pupils in TEEMUP schools made one additional month's progress in maths, on average, than those in the control group equivalent. This is our best estimate of impact, which has a high-security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes an effect of one month's less progress and an effect of up to two months' additional progress.

Secondary analyses found that pupils from both cohorts in TEEMUP schools made, on average, the equivalent of between two and five additional months' progress in cognitive, emotional, and behavioural self-regulation compared to pupils in control group schools, at the end of Reception. This suggests that TEEMUP had a positive impact on children's self-regulation outcomes, although there was uncertainty over whether these results could be sustained in the longer term.

The evaluation findings indicate that the more schools followed the TEEMUP requirements and teachers participated in the training, the better outcomes there were for pupils. However, there is some level of uncertainty around these results as only a limited number of schools met the full requirements of the programme.

The evaluation found that TEEMUP was very well received by teachers. Teachers who received TEEMUP reported that it had increased their confidence in their ability to teach maths, and that their practice had changed in line with TEEMUP objectives, with an increase in the use of games, books, and stories to teach maths. Changes also included working on maths home learning environment.

While TEEMUP largely worked as intended, some barriers were found to the delivery and implementation of TEEMUP. The impact of the COVID-19 pandemic on the education sector meant some delivery was performed online, schools had competing priorities, and staff turnover and absences were exacerbated. This resulted in workshop absences or, in some cases, withdrawal from the programme. One-third of Cohort 1 pupils did not move into a Year 1 class taught by a teacher at least 'minimally' trained in TEEMUP, as intended. As the evaluation found compliant schools saw better outcomes for their pupils, these factors could have limited the impact of the programme.

Cost

The estimated cost of TEEMUP is £16,538 per school over a three-year period, or £110 per child per year when averaged over three years. This figure is based on 50 pupils per school per year during trial delivery and includes costs for school recruitment, staff training, optional staff travel, optional staff cover, the purchase of optional additional materials and resources, and COVID-19 overheads.

Impact

Table 2: Summary of impact on primary outcome

Outcome / group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P-value	EEF cost rating
BAS3 ENC maths attainment / Cohort 1	0.07 (-0.05, 0.18)	1	🔒🔒🔒🔒🔒	1,330	0.26	£ £ £ £ £
BAS3 ENC maths attainment/ Cohort 1 FSM-eligible	0.13 (-0.10, 0.37)	2	N/A	238	0.27	N/A

BAS3=British Ability Scales, Third Edition; ENC=Early Number Concepts; N/A=not applicable.

Introduction

Study rationale and background

In England, the most recent Key Stage 2 assessments show that 27% of children do not meet the expected standard in maths at the end of primary school (Department for Education, 2023a). Longitudinal studies have demonstrated that early mathematical achievement is predictive of mathematics and general educational achievement in later life (Duncan *et al.,* 2007; Jordan *et al.,* 2009; Claessens, Engel, and Curran, 2014; Nguyen *et al.,* 2016), as well as socio-economic status across the lifespan (Ritchie and Bates, 2013). These findings highlight the importance of early numeracy education and the need for effective early years' interventions to narrow attainment gaps. Successful learning in early mathematics is linked to the development of self-regulation (Quigley, Muijs, and Stringer, 2021), which is recognised as a key competence to maintain lifelong learning (e.g. see Council of the European Union, 2002). Interventions that are designed to facilitate mathematical learning between the ages of three–five years old have been shown to have a strong positive effect on children's lives (Clements and Sarama, 2011). Indeed, interventions are considered the most cost-effective and efficacious if administered early in childhood (Easton and Gee, 2012; see EEF's Early Learning Toolkit relating to self-regulation strategies).

Studies have shown a disparity between the mathematics achievement of children from different economic backgrounds. Children from disadvantaged backgrounds start school behind their peers and are more likely to underachieve in maths across primary school (Jordan and Levine, 2009; Siegler, 2009; Department for Education, 2019; Hentges, Galla, and Wang, 2019). The quality of education provision impacts on children's outcomes, particularly for those from disadvantaged backgrounds (Melhuish *et al.,* 2015); thus, it is important to evaluate strategies to improve provision quality.

One way to improve mathematics provision, and thus outcomes of children, is to improve teachers' skills through continuing professional development (CPD) (Knowles and Fair Education Alliance, 2017). A previous efficacy cluster randomised controlled trial (RCT) known as Fostering Effective Early Learning (FEEL) (Melhuish *et al.,* 2016) involving 90 pre-school settings in Australia reported that implementing an evidence-based, structured CPD programme to early years' staff who taught four–five-year-olds over seven to nine months the year before they started school improved the quality of setting provision, as measured by the Early Childhood Environment Rating Scale – Extended (ECERS-E) and the Sustained Shared Thinking and Emotional Well-being (SSTEW) scale (Siraj *et al.,* 2018a). The trial also investigated 'indirect' impacts on child outcomes (secondary outcomes) shortly after the teacher training intervention had ended, and found evidence of improvements in early numeracy, verbal comprehension, and socio-emotional development among the children in the intervention group. These differences in child outcomes were not found to be statistically significant in alternative analyses that adjusted for baseline assessment scores.

This efficacy trial evaluated a modified version of the FEEL CPD programme, which has been developed for use with teachers of Reception and Year 1 schoolchildren in England with a focus on mathematics and self-regulation. This intervention is called Teaching Effective Early Mathematics and Understanding in Primary Schools (TEEMUP) professional development (PD) programme. A comparison of the FEEL and TEEMUP evaluations is provided in Appendix A Table 1.

An RCT, conducted and reported to the CONSORT (Consolidated Standards of Reporting Trials) standards (Cuschieri, 2019), is the best available design for answering questions of effectiveness (Cook and Campbell, 1979; Cook, Campbell, and Shadish, 2002; Torgerson and Torgerson, 2008), and was therefore, chosen to estimate the impact of the TEEMUP PD. Random allocation eliminates selection bias, and controls for all known and unknown variables, and other potential sources of bias in the TEEMUP RCT were minimised through design. For example, baseline assessments were conducted prior to school-level randomisation and outcome assessments were conducted blind to group allocation. An integrated implementation and process evaluation (IPE) was conducted to gain a comprehensive understanding of usual practice, compliance, and fidelity in intervention delivery, and to help explain any possible differences between randomised groups. An economic evaluation was also conducted to explore the cost of the intervention.

Intervention

The TEEMUP PD programme ran over a 17-month period, with the first five months consisting of nine teacher training sessions/workshops, subsequently supported by one mentor meeting per term thereafter (Summer Term 2022, Autumn Term 2022, and Spring Term 2023) with ad hoc mentor support by email/online meetings at the teacher's request, and a final workshop in Spring Term 2023. Nominated Reception and Year 1 teachers within participating schools received specialist training from the TEEMUP PD team in improving mathematics content/domain knowledge and how to support children's mathematics and self-regulation.

The PD aimed to allow teachers to:

- explore best practice in mathematics teaching;
- work together to support transitions into and across classrooms;
- effectively engage the children's home in their maths education;
- build their mathematical confidence, knowledge, and understanding;
- explore novel techniques to strengthen children's self-regulation; and
- effectively self-evaluate, plan for improvement, and monitor their own and the pupil's progress.

The primary goals of the TEEMUP PD were to:

- improve pupils' maths attainment at the end of Reception and Year 1, and
- improve pupils' personal, social, and emotional development (PSED) and self-regulation at the end of Reception and Year 1.

Further detailed information about the TEEMUP PD is included in the Template for Intervention Description and Replication (TIDieR) table (Table 3) and the logic model (Figure 1) below.

Table 3: Description of the programme using the TIDieR checklist

TIDieR item	Description
Brief name	Teaching Effective Early Mathematics and Understanding in Primary Schools (TEEMUP) professional development (PD) programme.
Why: Rationale, theory or goal of the elements essential to the programme	**Background** Mathematics skills are crucial for health, wealth, and quality of life (OECD, 2013; Muijs *et al.,* 2014; Nguyen *et al.,* 2016). Young children's mathematics achievements are known to be predictive of long-term educational attainment (Duncan *et al.,* 2007; Wylie and Hodgen, 2011; Claessens and Engel, 2013), and may be a better predictor of later life success than literacy (Bynner and Parsons, 1997; 2000). However, many children fail to acquire the mathematical skills necessary for success in adulthood, particularly those from disadvantaged backgrounds (Clements and Sarama, 2011). As they enter school some children are already behind their peers in mathematical skills. The home learning environment is an important predictor of children's mathematical development (Melhuish *et al.,* 2008; Walker, Shenker and Hoover-Dempsey, 2010). The type and frequency of maths activities that children engage with at home can influence their future mathematics performance. Further, the potential to improve children's mathematical skills is influenced by teachers' confidence and understanding in how to teach emergent and early mathematics. Teachers sometimes describe strong feelings of shame, humiliation, alienation, and disengagement, linked to their own school experiences with maths (Bibby, 1999; 2002). The potential of continuing professional development (CPD) is well documented as the most cost-effective way to maximise workforce capacities (e.g. Charalambous and Praetorius, 2018). The proposed CPD, designed to enhance the effectiveness of mathematics teaching, is in line with existing literature (Brophy, 1986; Muijs *et al.,* 2018), which suggests that effective teaching is likely to be a conglomerate of behaviours and that it is unlikely that one isolated behaviour will make the difference. Therefore, it is the combination of effective teaching behaviours that will lead to better mathematical performance in pupils. It is for this reason that the TEEMUP PD programme covers support for the

TIDieR item	Description
	teaching of mathematics, including problem-solving, thinking, and argumentation, as well as support for children's behaviour for learning (Ellis and Tod, 2015). **TEEMUP PD overview** In the TEEMUP study, improvement in children's mathematical abilities is the primary outcome, and there is a focus on developing children's mathematical understanding, reflecting early mathematics' strong associations with later success or failure in school (Claessens, Engel, and Curran, 2014). However, the TEEMUP PD is not confined to the mathematical knowledge, skills, and attitudes teachers require to support children's learning effectively; problem-solving, thinking, reasoning, and argumentation is strongly evident, which will build conceptual understanding in mathematics (Franke and Kazemi, 2001; Mulligan and Mitchelmore, 2009; Scharton, 2004). Additionally, aspects such as self-regulation, understanding disadvantage, support for cognitively challenging interactions, and intentional and relational pedagogy also feature (Whitebread, 2012; Ofsted, 2019). The evidence-based TEEMUP PD draws on current knowledge about effective mathematics teaching but is also designed to suit the participants and curricula requirements. It is informed by the delivery team's previous projects featuring CPD, which have been shown to shift classroom practice and enhance children's outcomes (Kingston, 2017; Siraj, Cheeseman, and Kingston, 2017; Siraj *et al.,* 2018a; 2018b).The TEEMUP PD was developed in line with the literature detailing the characteristics of effective PD (e.g. see Kingston, 2017) and the content, affect, and process of delivery are evidence-based (e.g. Siraj-Blatchford *et al.,* 2003; Halle, Metz, and Martinez-Beck, 2013; Pianta, 2012). Teachers are supported to make quality improvements through the use of practice development and self-reflection tools, building mathematical confidence, knowledge, and understanding. They are guided on effective self-evaluation, planning for improvement, and monitoring of their own and the children's progress. The training encourages active engagement and collaboration of the participants as they explore best practice in the teaching of mathematics and work together to support transitions and the home learning environment. The focus on theory to practice supports the development of the classroom climate and purposeful, developmentally appropriate mathematical instruction and behaviour for learning. Teachers learn from each other, the tutors—and the information and materials they provide—during the face-to-face sessions as well as when they visit them in school, taking the roles of coaches and mentors (Meyers, Durlak, and Wandersman, 2012; e.g. Siraj, Cheeseman, and Kingston, 2017; Siraj e*t al.,* 2018a; 2018b). The TEEMUP PD is designed to support the teaching of mathematics to all children. However, it is designed to support the kind of high-quality teaching, which includes assessment of the children's current understandings and achievements to ensure they are working within their zone of proximal development (Vygotsky, 1978); monitoring of the effectiveness of their teaching; active engagement of children with purposeful, real world, and interesting (to the children) problems; an emphasis on thinking, deep learning, and perseverance rather than achievements and getting the right answer; a classroom culture of trust, collaboration, and the belief that all children are mathematicians, which is likely to be particularly effective with children from disadvantaged backgrounds and those with Special Educational Needs (SEN) (National Council of Teachers of Mathematics, 2014). **TEEMUP PD and how it meets the characteristics of effective PD** *Content and affect* The content and affect (support for teachers' developing positive relationships with children and their families, and their beliefs and attitudes towards their teaching of mathematics) within the TEEMUP PD was developed following consideration of the literature on effective teaching, including the delivery teams own previous experiences of delivering PD. Generally, PD has been found especially useful to augment the knowledge and skills of teachers after initial training, it also keeps teachers up to date with research into best practice and effective teaching. In this study, effective teaching is considered to be in classes where the teachers enhance and support children's learning outcomes. Coe *et al.* (2014) suggested the following features are important for teachers to be effective in schools. 1. Pedagogical content knowledge Effective teachers have a good knowledge of the subject they are teaching, and the concepts that children must grasp to understand what they intend them to learn. Teachers need to be aware of the pupil's responses to the learning, and how they are thinking about the content, including being able to identify pupil's misconceptions. 2. Quality of instruction The quality of interactions or sustained shared thinking (see Siraj-Blatchford *et al.,* 2002) the teacher has with, and supports between, the pupils is fundamental to the quality of instruction. Elements of practice, such as reviewing previous learning, effective questioning, providing models, using formative assessment, giving adequate time and practice to embed skills securely, and progressively introducing new learning (scaffolding), are also prevalent within high-quality instruction.

TIDieR item	Description
	3. Classroom climate The expectations set within the classroom and the relationships built are important for effective teaching and learning. When teachers are positive and respectful, setting learning within the pupil's Zone of Proximal Development (Vygotsky, 1978) that is challenging but also attainable, it builds and supports the pupil's sense of self-worth. When the teacher values effort and persistence over ability, they build pupil's resilience and support their developing self-regulation. 4. Classroom management A teacher's abilities to support pupil's behaviour for learning through the classroom climate (see above) and how they structure and organise their learning—for example, by avoiding long periods of waiting, making efficient use of learning time, coordinating and making resources and space accessible, and promoting behaviour for learning through the use of clear and consistent rules—are all relevant to effective teaching. These environmental factors set the scene for good learning, enabling high quality of instruction. 5. Teacher beliefs Teachers' theories about learning and how it happens to impact on the practices they adopt in their classrooms. They impact on the role they take as the teacher, as well as the expectations they have of the pupils and what they aim to achieve. Teachers who believe pupils capable of co-constructing knowledge, rather than seeing learning as reliant solely on the teacher and on rote learning techniques, support pupils to be creative thinkers and problem-solvers. 6. Professional behaviours While these may be more indirect aspects of effective teaching, there is some evidence to suggest that behaviours exhibited by teachers, such as reflecting on and developing professional practice, including participation in PD and collaborating with and supporting colleagues, can impact on pupil's outcomes. In addition, liaising and communicating with parents, including supporting the home learning environment (particularly with younger children) can also support children's learning. *Process of delivery* The process of delivery was also influenced by the literature on the effective characteristics of PD. For example, teachers are: encouraged to collaborate; specific teaching (during face-to-face sessions) is combined with follow-up coaching and mentoring in the classrooms; attention is paid to the intensity, and duration of the TEEMUP PD, and attendance will be monitored; at least two members of staff from each school will join the TEEMUP PD and a member of the senior management team (SMT) will act as the main contact; and, finally, networks of connections across and within the schools will be encouraged to support sustainability and the notion of lifelong learning. The goals of the TEEMUP PD are to: • Improve pupils' maths attainment at the end of Reception and Year 1. • Improve pupils' personal, social and emotional development (PSED) and self-regulation at the end of Reception and Year 1. • Improve attainment in maths, PSED, and self-regulation for children from disadvantaged backgrounds. • Improve staff understanding and implementation of Developmentally Appropriate Practice, in particular with maths instruction. • Improve staff capacity to relate mathematics to the rest of the curriculum. • Improve partnership working with families on home learning environment. Longer-term goal: Improve SATS results at the end of Key Stage 2 (although not investigated within the context of this specific trial).
Who: Recipients of TEEMUP PD	A minimum of two teachers, one from Reception and one from Year 1 (three teachers are welcomed) from state-funded primary and infant schools, primarily located in the East of England. In addition, one member of the SMT, who may attend Days 1 and 2, to support the teachers' in implementing changes in school. No cascading from trained teachers to teachers beyond their classroom is required/expected.
What: Physical or informational materials used in the programme	Teachers will receive specialist training from the Oxford PD team in how to support children's mathematics and self-regulation: • The face-to-face training originally consisted of two full days (9.30 to 16.00) in Spring Term 2022 and eight half-days (14.00 to 17.45) face-to-face workshops (seven half-days in Spring Term 2022 and one half-day in Summer Term 2023). This was revised to comprise two full days (9.30 to 16.00) in Spring Term 2022 and eight half-days 13.30 to 16.50 face-to-face workshops (seven half-days in Spring Term 2022 and one half-day in Summer Term 2023). The timings of the workshops were changed and reduced after consultation with schools to reduce travel time after work hours and reduce stress on teachers, which were deemed important factors given COVID-19. The two consecutive days were followed by seven half-day sessions run once a fortnight, allowing time between sessions to use the new ideas and activities and involve other staff within their team. The final, half-day, follow-up workshop was offered in 2023. Again, in response to COVID-19 and the resulting staffing issues—including illness, additional stress, staff retention difficulties, finding cover to

TIDieR item	Description
	release staff to travel to and attend face-to-face training sessions, and concerns over mixing with teachers from other schools and spreading infections—a lighter touch training of all of the PD was also offered online. At roughly the same time as the face-to-face training, 11 sessions were run online. Each session was run twice on different days during a week (Day 1 and Day 2 involved four online sessions and then each of the seven workshops was presented separately). Teachers were asked to do some preparation before sessions, including watching online video excerpts and completing some activities. In the second year of the study in the first Autumn Term of the new academic year, 12 online PD sessions were run for staff new to TEEMUP due to staff changes (many changes were linked to COVID, e.g. long COVID, stress linked staff turnover, leaving the profession, and maternity leave). Ten online PD sessions were run once, with the expectation that staff would spend additional time preparing for the sessions, and two individual mentoring online sessions were offered to each participant. • **Specialist needs-based coaching/mentoring in schools.** Following the workshops (from Summer Term 2022 to Summer Term 2023), mentoring/coaching was provided on a needs-based model, with a minimum of three coaching/mentoring sessions. These sessions were designed to support the implementation of changes, adapting the approaches to suit the school's context and children/families, and getting other staff (e.g. teaching assistants) involved. Where schoolteachers required more support and asked for it during meetings, the mentors offered further support. In addition, one additional school visit for headteachers and SMTs was offered to all schools to discuss the TEEMUP process and how to support participants. This was because many SMTs found they could not be released to attend the PD, due to staffing pressures linked to COVID-19. During the second year, the online training included two individual online mentoring sessions. • **Website with PD resources and additional materials.** A dedicated website by the delivery team supported in-class teaching, providing additional materials, cascading the approach to other staff, and provides information and ideas to support early maths and self-regulation learning at home.
What: Procedures, activities and/or processes used in the PD	**Vehicles for delivering the PD (How)** Original: A minimum of 53 hours of face-to-face workshops (two full days at 6 hours 30 minutes each day; seven half-day workshops and one later half-day workshop i.e. eight half-days at 3 hours 45 minutes each) and in-school mentoring (a minimum of three visits at 3 hours 30 minutes each). Revised: A minimum of 50 hours of face-to-face workshops (two full days at 6 hours 30 minutes each day; seven half-day workshops and one later half-day workshop i.e. eight half-days at 3 hours 20 minutes each) and in-school mentoring (a minimum of three visits at 3 hours 30 minutes each). The reduction in time was agreed with headteachers/SMT and designed to accommodate teachers, during the COVID-19 pandemic. The workshops included the use of PowerPoints, videos, mathematical materials, examples of good practice, assessment tools for use at the classroom and child-level, games, and books. The sessions were held every two to four weeks, and were sufficiently spread out to allow for practices/strategies to be trialled in between them. The mentoring visits were made to individual schools. Also, in response to COVID-19 and associated staffing issues, alongside the face-to-face PD, a lighter touch set of online sessions were offered. They consisted of 16 and a half hours online plus at least three, three and a half hours face-to-face mentoring sessions (totalling 27 hours). These sessions were run approximately fortnightly following the same pattern as the face-to-face sessions. While this added to the teaching burden of the tutors and mentors, only two teachers accessed the training totally online in the first year. Other teachers took a blended approach, accessing online sessions only when they missed face-to-face sessions, due to illness or lack of cover, for example. In the first term of the second academic year, the entire PD was offered online again. It consisted of 18 and a half hours online followed by at least two, three and a half hours of face-to-face mentoring sessions (totalling 25 hours). Each online session had preparation work to be completed prior to attending the session. These sessions were run weekly in the first school term of the second year of the study. The sessions were offered weekly rather than fortnightly in order to upskill the new teachers on the TEEMUP PD as rapidly as possible. While this meant participants became familiar with the materials quickly, it did not allow much time to use the approaches, materials, and ideas between the sessions. Participants were given access to a bespoke online learning/knowledge base website with discussion forums, materials, articles, suggestions of games activities, resources to download, and copies of the PowerPoints and videos. They were also offered ongoing access to TEEMUP mentors/trainers through workshops, school visits, Skype™, email, and phone. **Specific content of the PD (What)** • Two TEEMUP reflective self-assessment scales were introduced early and used throughout the PD: i) Behaviour for Learning scale; and ii) Improving Maths Practice scale for Reception and Year 1. • A set of ideas/activities and resources to try in between sessions (from theory to practice) in the classroom. Together with a reflective practice and planning framework designed to support planning for change, capture changes, support sharing of ideas/changes with others, and evaluate any impact those changes made within classes.

TIDieR item	Description
	• A set of maths materials/information designed to support planning, implementation, and evaluation of activities and lessons to support pupil's learning. • Formative assessment ideas and examples, for measuring mathematical achievements to monitor child progress and inform planning. • A set of materials, ideas, and games designed to support teachers working with parents/carers on the home learning environment. • A set of ideas/resources for use in class to support pupil's behaviour for learning. • The opportunity to work in collaboration with other teachers from their own school. While it was originally hoped that teachers would also work with teachers from other schools during the workshops, supporting the development of a community of learners, this was not possible due to COVID-19. The teachers sat on tables with colleagues from their own schools and were kept separate from teachers from other schools throughout the sessions. This was an agreement made with headteachers and the SMT prior to starting the PD as it was hoped it would reduce the possibility of cross infection. Initially, many of the workshops were meant to include a carousel of practical ideas/activities, however this was not possible given the restricted movements needed to keep teachers from different schools separate from each other. The face-to-face sessions: **Day 1** Introduction to the TEEMUP study and what to expect Session 1: All about maths; why maths is important, some common myths about maths, maths anxiety, charting maths in my classroom Session 2: What is maths like in my classroom? This involved use of the self-reflection scale – Improving Maths Practice, knowledgebase, and quality improvement proforma **Day 2** Session 1: Improving Maths Practice Session 2: Supporting self-regulation and Behaviour for Learning **Half-days** Session 1: Early number sense and subitising Session 2: Counting and cardinality Session 3: Number relationships Session 4: Addition and subtraction Session 5: Reasoning and problem-solving Session 6: Mathematical talk, communication, and collaboration. Session 7: Behaviour for learning Mentoring/coaching: A minimum of three visits per school Session 8: in Summer Term 2023 TEEMUP practice, implementation, and impact evaluation day
Who: **PD providers/ implementers**	Professor Iram Siraj, Dr Denise Kingston, Stephanie Cottriall, and Steph Flower acted as tutors and mentors. Technical and website support was provided by Judy Barrett.
How? **Mode of delivery**	PD workshops for teachers were provided in local centres/schools in groups of 12 to 30 with an option of attending another session in an adjacent area if a session was missed or viewing a presentation of the session on the website or attending an online session. There were six groups for all the face-to-face sessions, held in different geographical areas. The half-day sessions ran on different days once a fortnight to allow participants to attend different sessions if possible and needed. Originally, only four hubs were planned for, but difficulty in recruiting during COVID-19 required expanding the areas and increase the number of hubs to six. PD face-to-face sessions were videoed so that they could be viewed at a later date by those who missed sessions due to illness or for other reasons. In addition, two online sessions (available on different week days) for each of the ten sessions (each lasting for one and a half hours) were available for staff who missed face-to-face sessions. A minimum of three sessions of mentoring and coaching were offered to all schools i.e. small group and/or one-on-one support as needed/appropriate. This was agreed with the schools. Additional support was provided via email, video, or telephone calls where required. Contingency planning to respond to non-attendance due to, e.g. illness, external circumstances, or for newly appointed staff, resulted in light-touch online sessions being made available alongside the face-to-face sessions in the first year of the study and another set of online PD sessions being made available at the beginning of the second year of the study for staff new to TEEMUP.

TIDieR item	Description
Where? Location of Delivery	The programme was available in the East of England and surrounding areas. Face-to-face sessions were located so that they were within an hour or less travel time for most schools. However, for those schools slightly outside the Eastern region the distance may have been greater, with a maximum travel time of one and a half hours. The increases in distance travelled reflected the difficulty in recruiting schools during COVID-19.
When and how much? Duration	Schools were supported for approximately 16 to 17 months. Initial TEEMUP PD consisted of five months of nine initial group training workshops, followed by 12 months of mentoring support (generally one visit per term) and a final workshop in Spring Term 2023. For new teachers joining in September 2022, they had three months of group online training workshops followed by six months of mentoring support (generally one visit per term).
Tailoring? Adaptation of the programme	The workshops were standardised and the delivery team followed a manual. However, there were some individual differences, as part of the sessions, the trainers were responsive to feedback from schools/cohort of teachers regarding their changes and practice between sessions. Teachers planned their own activities/lessons as practice between sessions, adapting the materials, strategies, and ideas in the sessions to suit their context. The mentoring and coaching approaches were also responsive to needs of the staff and school context.
How well (planned): Strategies to maximise effective implementation	In addition to the extensive training, resources, and support outlined in the sections above, the following strategies were employed to maximise effective implementation: • Piloting of newly developed materials prior to PD (through expert panel review of resources and materials and a pilot with at least four schools). • The introductory two days took teachers step-by-step through the process and familiarised them with the online website, the self-reflection tools, and what to expect throughout the PD. • Mentors phoned participants, if any concerns arose, rather than emailing them. • Face-to-face visits to schools were carried out to support implementation and to review how revised approaches and resources were working in practice. The mentors collected some photographic evidence/short videos to support the last workshop of the PD. • Additional resources and materials for use in class and for sharing with families were available on the learning platform/website. • Teachers who could not attend a session were invited to an alternative one or directed to an online session and then a tutor followed up with them. • Additional online training was given where needed, e.g. change of teacher/new teacher. • A whole-school staff development session explaining the study (explaining how trainers/mentors and teachers worked together) was offered as an online resource for staff to use in schools if they wished. • An early information session to gain school 'buy-in' was given to headteachers, senior leaders, and maths leads, run by mentors/trainers. • In addition to the planned face-to-face sessions at the start of the study, two sets of online PD were made available. One set ran alongside the face-to-face sessions and another set ran at the beginning of the school year in the second year of the study.

Source: Written by Kingston, Siraj, Melhuish, and Barrett 12 May 2021.

Changes to the intervention due to the COVID-19 pandemic

As detailed in Table 3, the intervention offered participating Reception and Year 1 teachers 50 hours of face-to-face workshops:

- two consecutive full days, 9:30 a.m. to 4:00 p.m. (six and a half hours per day) delivered in January 2022; and

- seven half-day training workshops, 1:30 p.m. to 4:50 p.m. (three hours and twenty minutes each) delivered every two to four weeks between January to June 2022 (these were originally intended to be 40 minutes longer).

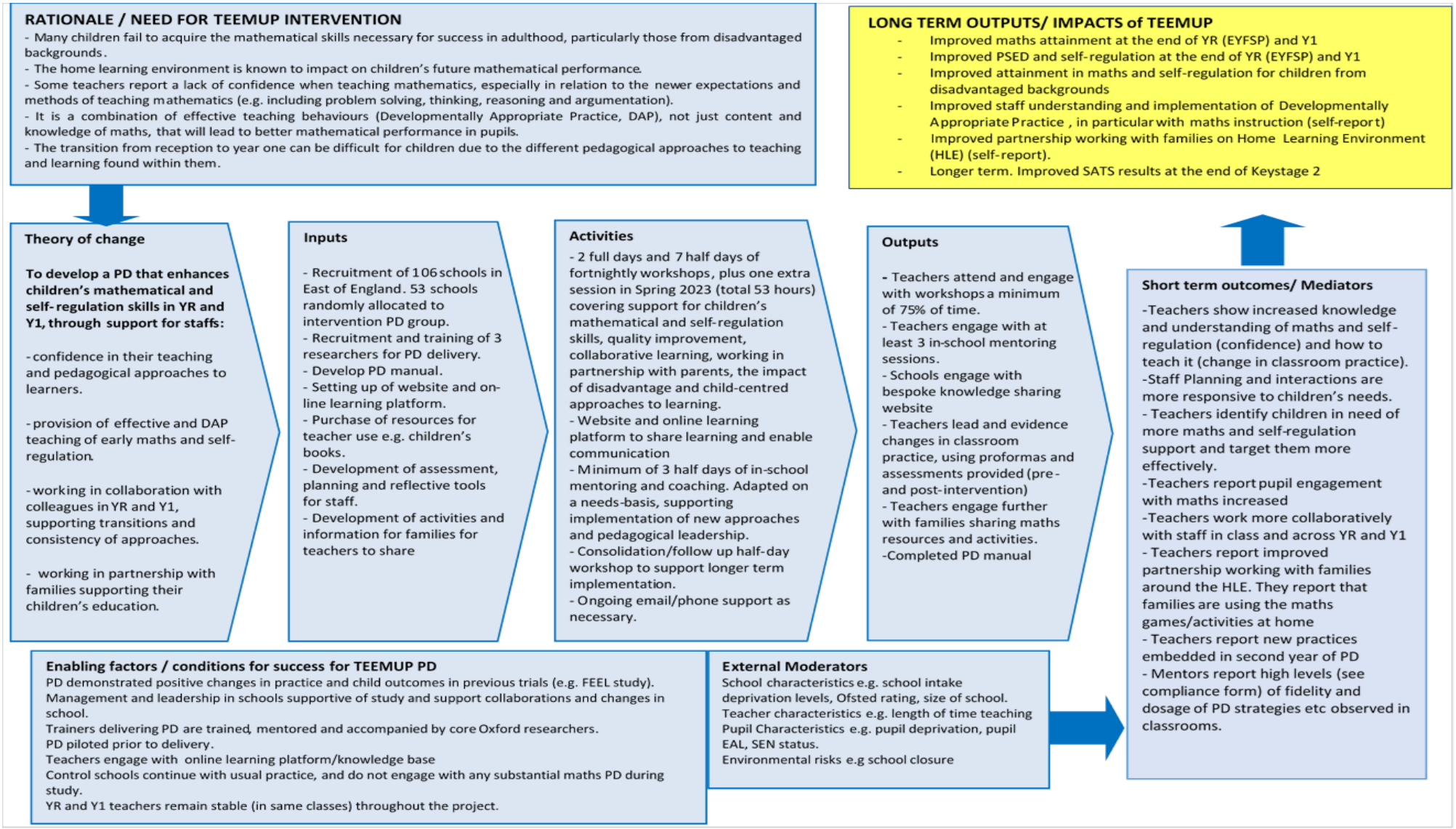
One half-day training workshop, 9:30 a.m. to 2:30 p.m. (five hours) in April 2023 to May 2023. The workshops were based at six hubs located in Milton Keynes, Oxford, Bury St Edmunds, Norwich, Barnett, and Peterborough. The delivery team initially intended that there would be four hubs but increased this to facilitate school recruitment and reduce participating teachers' travel time. Teachers were allocated to their hubs by the delivery team; however, there was flexibility for a teacher to change hubs should one be closer to their home. Social distancing requirements up until April 2022 meant teachers attending the hubs had to be seated in 'bubbles' (seated with only members of their school), which reduced collaboration with teachers from other schools; this was a change to the original TEEMUP PD.

Due to COVID-19 restrictions and the impact that COVID-19 had on staff absence and staff retention within schools in the 2021–2022 academic year, the delivery team were required to host training online via Zoom to some schools/teachers who were unable to attend training at a hub. This was a necessary response to environmental factors at the time, which were out of the team's control. Delivering the TEEMUP PD online was a necessary adaptation to prevent higher levels of school attrition. The delivery team considered the training delivered online during this period to be a lighter touch in comparison to that delivered at a venue.

During 2021–2022, teachers unable to attend the first two consecutive full-day training workshops were offered the training online (preparation work together with online sessions lasted one full day [six hours] followed by two sessions each lasting two and a half hours). The subsequent workshops delivered online were offered to schools on two occasions in a week, roughly a fortnight after the corresponding face-to-face sessions. To facilitate online sessions, video links, and activities, materials were sent to teachers beforehand, and they were asked to watch the videos/complete some activities prior to attending the sessions. The online workshops lasted one and a half hours. Teachers could also watch recorded training workshops if they were unable to attend both the venue or the online training.

To accommodate staff changes within schools **in the second year of the trial**, any newly nominated Reception and/or Year 1 teachers were approached by the delivery team and offered the PD online. This involved ten weekly online sessions between 3:00 p.m. to 4:30 p.m. and two online mentoring sessions (all of these online sessions lasted for one and a half hours). These sessions ran weekly from September 2022 to December 2022. As the online sessions were held weekly, it meant teachers did not have time to try out new approaches in their classrooms prior to the next session. Nevertheless, the frequency of the weekly sessions was considered necessary to ensure teachers had as much TEEMUP training as possible before the end of the intervention delivery period. The delivery team acknowledged that these sessions were not as in-depth as the workshops due to time restrictions. Teachers who received such online training also did not have time to receive the full number of intended mentor school visits.

Figure 1: TEEMUP logic model



Evaluation objectives

The latest version of the protocol and Statistical Analysis Plan (Baird and Fairhurst, nd) for the trial can be found on the EEF project website.

Impact evaluation research questions

Primary outcome

What is the impact of the TEEMUP PD, in comparison to usual teaching practice, on children's maths attainment at the end of Year 1? [Cohort 1 only]

Secondary research questions

Cohort 1
In comparison to usual teaching practice, what is the impact of the TEEMUP PD at the end of Reception *and* Year 1 on children's self-regulation and PSED (measured using the Child Self-Regulation and Behaviour Questionnaire [CSBQ])?

Cohort 2
Included to investigate the impact of a Reception teacher at the end of receiving TEEMUP PD on children's outcomes (exploring teacher development/experience and allowing time for practice to embed).

What is the impact of the TEEMUP PD, in comparison to usual teaching practice:

- on children's maths attainment at the end of Reception?
- on children's self-regulation and PSED (measured using the CSBQ)?

Both cohorts

What is the impact of the TEEMUP PD, in comparison to usual teaching practice, on children's Early Years Foundation Stage Profile (EYFSP) scores at the end of Reception, including Mathematics Early Learning Goals (ELGs), self-regulation, PSED, ELGs, and general development?

Is the TEEMUP PD effective in raising the maths attainment of children who are eligible for FSM at the end of Reception (Cohort 2) and Year 1 (Cohort 1), in comparison to usual teaching practice?

Teachers

Is the TEEMUP PD effective in improving nominated teachers' confidence in supporting children's maths development in comparison to usual teaching practice?

IPE research questions

1. Is fidelity to the TEEMUP PD being observed?
    - 1.1 Are participating teachers attending/accessing the available training?
    - 1.2 Are the different components of the TEEMUP PD materials and resources (e.g. reflective self-assessment scales, planning framework, use of formative assessments) being used as intended?
    - 1.3 Are participating schools engaging with the school's allocated TEEMUP PD mentor as expected?
    - 1.4 What are the barriers and/or facilitators to teachers engaging with the TEEMUP PD training and the mentor?
    - 1.5 What constitutes necessary conditions (enabling factors) for participating teachers and schools to engage with the intervention as intended?
    - 1.6 Do outcomes vary in line with compliance?
2. To what extent is the TEEMUP PD implemented as planned within schools?
    - 2.1 To what extent do teachers implement the TEEMUP PD in their teaching practice?

2.2 Have teachers adapted the intervention to make it more suitable for them, if so, how?
2.3 What are the facilitators and/or barriers to teachers implementing the TEEMUP PD?
2.4 What are the necessary conditions for teachers to implement TEEMUP PD into practice?

3. What are different stakeholder viewpoints of the TEEMUP PD?
   3.1 What are teachers' perceptions on the usefulness and quality of the intervention as a whole and its components e.g. training, mentor support/visits, maths practice, and Behaviour for Learning scales, resources/materials?
   3.2 What is the perceived impact of the TEEMUP PD on teacher's maths practice, teachers' confidence in teaching children maths, and teachers' confidence in their own maths abilities?
   3.3 What is the perceived impact of the TEEMUP PD on children's maths outcomes and self-regulation? Are there any perceived differential intervention benefits among disadvantaged children?
   3.4 How can the TEEMUP PD programme be improved?
   3.5 What is the perceived impact of the TEEMUP PD on teachers' practice and confidence in relation to children's self-regulation?
4. What is 'usual practice' in all schools and has this changed in schools that have received the intervention?
   4.1 What is teacher's usual maths practice and has this changed in schools after receiving the TEEMUP PD?
   4.2 How do teachers usually engage with children's families and the home learning environment and has this changed as a result of the TEEMUP PD?
   4.3 How, if at all, do teachers work collaboratively within and across schools, and has this changed as a result of the TEEMUP PD?
   4.4 How frequently, if at all, do teachers use formative feedback within their usual practice, and has this changed in schools that received the TEEMUP PD?
   4.5 How frequently, if at all, do teachers use specific tools to reflect on their practice, and has this changed in schools that received the TEEMUP PD?
   4.6 What is usual practice in relation to the transition between Reception and Year 1, and has this changed in schools that received the TEEMUP PD?
5. To what extent does the TEEMUP impact evaluation adhere to the proposed plan?
   5.1 Does the child recruitment and assessment process adhere to the plans proposed in the protocol?
   5.2 Are there any sample attrition effects? If so, how might that affect the estimates of the impact of the TEEMUP PD?
6. Is each stage of the intervention logic model supported by evidence and learning from the IPE?
   6.1 On review, after experience of PD delivery, do the delivery team consider any changes to the logic model necessary?
7. What can be learned from the efficacy trial to inform an effectiveness trial?

Ethics and trial registration

The University of York, Health Sciences Research Governance Committee granted ethical approval for the trial on 18 November 2020 and the University of Oxford, Department of Education Research Ethics Committee on 13 January 2020.

A Memorandum of Understanding (MoU) was signed by all schools to cover the requirements of the project (see Technical Notes). Participating schools also signed a data sharing agreement issued by the University of York (see Technical Notes).

The trial's International Standard Randomised Controlled Trial Number (ISRCTN) Registration is: 25478558, and available to view here: https://doi.org/10.1186/ISRCTN25478558.

Data protection

All data collected for the trial was treated with the strictest confidence and processed and stored in compliance with the General Data Protection Regulation (GDPR) and Data Protection Act, 2018 (GDPR, 2016). The University of York was the data controller and a data processor, as defined in the GDPR. Once the data has been submitted to the Office for National Statistics (ONS) Secure Research Service (SRS) for archiving in the Education Endowment Foundation (EEF) data archive, the EEF will hold data controller responsibility.

In line with the University of York's charter, which states that the University advances learning and knowledge by teaching and research, personal data was processed under Article 6 (1) (I) of the GDPR ('processing is necessary for the performance of a task carried out in the public interest') and Special Category data under Article 9 (2) (j) ('processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes'*)* of the GDPR (GDPR, 2016).

Schools were provided with information regarding data protection and data sharing for the trial within the MoU and the data sharing agreement (issued by the University of York). As outlined in detail in the 'School and participant selection' section, parents/carers were provided with an information sheet about the trial via schools and had the opportunity to withdraw their child from the evaluation elements of the project (data sharing and assessments) before these started.

A unique trial identification (ID) number was generated for each pupil when their details were entered into the trial management system. The trial management systems and trial data (impact evaluation and IPE data) were held on secure University of York servers with access limited to specified staff members of the University of York.

Electronic transfers of personal pupil data between schools and the York Trials Unit, University of York were completed via encrypted spreadsheets sent through the University of York DropOff facility (a secure file transfer service). Staff eConsent forms (to participate in interviews) and survey data were collected using Qualtrics survey software.

The dataset for statistical analysis was pseudonymised. No schools, staff members, or pupils are identifiable in the report nor will be identifiable in the dissemination of any results. Electronic data and paper documents including identifiable personal pupil data will be securely archived and then disposed of by the York Trials Unit five years after the publication of the final report (in 2030). Any identifiable personal data about adult data subjects (e.g. school staff) will be kept for five years after the publication of the final report (in 2029). Anonymised electronic data and paper documents will be kept indefinitely by the evaluation team and may potentially be shared with other research teams or archiving organisations.

Project team

Evaluation team

The evaluation team was based at the York Trials Unit, University of York:

- **Dr Lyn Robinson-Smith, Assistant Professor in Health and Education Research.** Experienced in leading and delivering large trials, particularly in the early years. Joint principal investigator and responsible for the design of evaluation, had oversight and management of the trial and led the writing of the final report. On maternity leave from February 2024 to December 2025.

- **Caroline Fairhurst, Senior Statistician.** Currently supports a number of trials, including several trials funded by the EEF, within the York Trials Unit. Contributed to the overall design, undertook the statistical analysis and cost evaluation, and will take responsibility for archiving data with the FFT. From August 2022, was joint principal investigator having oversight of the impact evaluation and the overall trial. On maternity leave from April 2024 to March 2025.

- **Hannah Ainsworth**, **Education and Health Care Trial Manager.** Joint principal investigator, led the impact evaluation and had oversight of the trial until August 2022.

- **Professor Carole Torgerson.** Expert in RCT design and conduct and has been the principal investigator or a co-investigator on over 25 RCTs. Contributed to the overall design, conduct, and report of the evaluation.

- **Professor David Torgerson, Director of York Trials Unit.** Worked on numerous RCTs, including many in education and the social sciences. Supported the design and conduct of the trial and undertook corresponding author duties for the final report from April 2024.

- **Kalpita Baird (née Joshi), Statistician.** Currently supports a number of trials within the York Trials Unit. Undertook the randomisation.

- **Dr Katie Whiteside.** Worked on a number of RCTs evaluating education and health care interventions. Contributed to trial coordination, data management, and to writing the final report.

- **Jess Hugill-Jones.** Worked on a number of RCTs evaluating education and health care interventions. Has a background in psychology and is an experienced primary school teacher. Was a trial coordinator for the evaluation from March 2020 to July 2021, and again from March 2023. Contributed towards IPE data analysis and report writing.

- **Louise Elliott.** Worked on a number of RCTs evaluating interventions. Was the trial coordinator for the evaluation and managed the trial data between June 2021 and August 2022.

- **Dr Heather Leggett, Applied Researcher, York Trials Unit.** Has experience of conducting quantitative and qualitative research across a range of projects in public health and education. Contributed to designing the IPE, and conducted some IPE data collection and analysis elements.

- **Imogen Fountain, Trial Support Officer.** Trial support for the evaluation and has experience working on numerous trials by the EEF.

- **Emma Standley.** Completed some of the IPE data collection.

- **Dr Kerry Bell**, **Economist and Trial Manager.** Substantial experience of trials within an educational context. Was involved in the design of the cost evaluation.

- **Tom Davill, Trial Support Officer.** Trial support for the evaluation and assisted in delivering outcome testing.

- **Dr Rachel Bottomley-Wise (née Carr).** Has a background in health psychology and has experience in varied trials, including those involving children and parents, and health behaviours during the postpartum period. Between September 2022 and May 2023, undertook some trial coordination duties including recruiting research assistants for post-testing and conducted several IPE interviews.

Delivery team

- **Professor Iram Siraj, Department of Education, University of Oxford.** Expert in early childhood and primary education. Principal investigator of the TEEMUP study and has led a number of longitudinal studies and RCTs looking at the effects of early education and interventions on long-term developmental outcomes. Leads the delivery team and worked on the development of the TEEMUP PD, the TIDieR table, the logic model, and complier average causal effect (CACE) analyses criteria.

- **Professor Edward Melhuish, Department of Education, University of Oxford.** Expert on longitudinal studies, child development, and evaluation of interventions. Co-investigator on the TEEMUP study. Previous work has included longitudinal studies, and evaluation studies that have had an impact on policy in the UK and other countries. Consultant to the World Health Organization, the European Union, the Organisation for Economic Co-operation and Development (OECD), and several research councils internationally.

- **Dr Denise Kingston, Department of Education, University of Oxford.** Specialist in children's development, socialisation, and inclusion, with a background in educational psychology. Current interests and research focus on effective PD and early childhood pedagogy and practice, including early mathematics. Senior researcher, project manager, and co-investigator on the TEEMUP study.

- **Judy Barrett, Department of Education, University of Oxford.** The delivery team administration manager and leads on the liaison and recruitment of schools, the development of the website, and administration and support for the delivery team and delivery team data.

Methods

Impact evaluation design

The impact evaluation consisted of a two-armed, cluster RCT. Randomisation was at the school level with schools allocated to the intervention group receiving the TEEMUP PD and schools in the control group continuing with 'business as usual' (Table 4). Two cohorts of pupils were recruited to take part in the evaluation: children beginning Reception in September 2021 (Cohort 1); and children beginning Reception in September 2022 (Cohort 2). Cohort 1 was followed up until the end of Year 1 (June 2023 to July 2023). Cohort 2 was followed up until the end of Reception (June 2023 to July 2023).

At the outset of the evaluation, before randomisation, schools nominated a minimum of two teachers, one from Reception and one from Year 1 (but up to three teachers were welcome), to participate in the TEEMUP PD if their school was allocated to the intervention group. Teachers in intervention schools were invited to receive the TEEMUP PD and associated support over a 17-month period (January 2022 to May 2023). Changes to practice would be expected to build over this period and therefore, the evaluation sought to investigate the impact of the TEEMUP PD on pupils in Cohort 1 (who, at the end of Year 1, should have been taught by Reception and Year 1 teachers receiving TEEMUP PD) and Cohort 2 (who, at the end of Reception, should have been taught by the nominated Reception teachers at the end of receiving the full TEEMUP PD). Participating schools/teachers signed an MoU, which indicated their agreement to, wherever possible: i) retain nominated Reception and Year 1 teachers in their respective year groups for the duration of the trial; and ii) keep participating pupils in classes taught by nominated Reception and Year 1 teachers.

All participating schools (both the intervention and control group) received £250 following recruitment and baseline data collection of Cohort 1, and £500 following completion of the British Ability Scales, Third Edition (BAS3) Early Number Concepts (ENC) post-test data for Cohorts 1 and 2 in June 2023 and July 2023. These payments were pre-specified in the MoU document signed by schools at the start of their participation in the trial.

Due to the COVID-19 pandemic, the timeline of the efficacy trial was amended within the protocol (Ainsworth *et al.,* 2023).

Table 4: Evaluation design

Trial design, including number of arms		Two-armed, cluster RCT, two cohorts Cohort 1 followed for two years: Reception 2021–2022 to Year 1 2022–2023 Cohort 2 followed for one year: Reception 2022–2023
Unit of randomisation		Primary schools
Stratification variable (s) (if applicable)		Minimisation was undertaken to ensure balance across the trial arms on: the percentage of pupils eligible (ever-6) for FSM in the school (latest available data), the percentage of pupils identified as having English as an Additional Language (EAL) in the school (latest available data), and the school's geographical location
Primary outcome	Variable	Maths attainment at end of Year 1 (Cohort 1 only)
	Measure (instrument, scale, source)	BAS3 ENC, score range of 0–35, GL Assessment. Collected by blinded evaluation team research assistants
Secondary outcome(s)	Variable(s)	**Cohort 1** (Reception 2021–2022, Year 1 2022–2023) PSED at end of Reception and end of Year 1 Self-regulation at end of Reception and end of Year 1 Routinely collected maths, self-regulation, PSED, and general attainment at end of Reception **Cohort 2** (Reception 2022–2023)

		Maths attainment at end of Reception PSED at end of Reception Self-regulation at end of Reception Routinely collected maths, self-regulation, PSED, and general attainment at end of Reception **Teachers** Teacher confidence (in teaching children maths), during intervention and at the end of intervention (Reception and Year 1 teachers)
	Measure(s) (instrument, scale, source)	**Maths attainment** BAS3 ENC, score range of 0–35, GL Assessment. Collected by blinded evaluation team research assistants **PSED** CSBQ, 34-items yielding seven subscales: Sociability; Prosocial behaviour; Externalising problems; Internalising problems; Cognitive self-regulation; Emotional self-regulation; and Behavioural self-regulation. Collected by nominated Reception and Year 1 teachers **Self-regulation** Using the three self-regulation subscales (Cognitive, Behavioural, and Emotional) of the CSBQ (17 items). Collected by nominated Reception and Year 1 teachers **Routinely collected maths, self-regulation, PSED, and general attainment at the end of Reception** EYFSP data collected by teachers at the end of Reception accessed from the National Pupil Database (NPD) • Maths (Number ELG and Numerical Patterns ELG, combined) • Self-regulation ELG PSED (Self-regulation ELG, Managing self ELG, and Building relationships ELG combined) • General Attainment (All 17 EYFSP ELGs average total point score and whether Good Level of Development has been 'met') **Teacher confidence: Maths** Adapted 'Early Math Beliefs and Confidence Survey' by Chen *et al.* (2014). One subscale: 'Confidence in Helping Children Aged 4–6 Learn Maths'
Baseline for primary outcome	Variable	Maths attainment at start of Reception November 2021 to January 2022 (Cohort 1 only)
	Measure (instrument, scale, source)	BAS3 ENC, score range of 0–35, GL Assessment. Collected by blinded evaluation team research assistants
Baseline for secondary outcome(s)	Variable	**Cohort 1** (Reception 2021–2022, Year 1 2022–2023) PSED at start of Reception Self-regulation at start of Reception **Cohort 2** (Reception 2022–2023) PSED at start of Reception Self-regulation at start of Reception **Teachers** Teacher confidence (in teaching children maths) at baseline (Reception and Year 1 teachers)
	Measure (instrument, scale, source)	**PSED** CSBQ, 34-items yielding seven subscales: Sociability; Prosocial behaviour; Externalising problems; Internalising problems; Cognitive self-regulation; Emotional self-regulation; and Behavioural self-regulation. Collected by nominated Reception and Year 1 teachers

		Self-regulation Using the three self-regulation subscales (Cognitive, Behavioural, and Emotional) of the CSBQ (17 items). Collected by nominated Reception and Year 1 teachers **Teacher confidence and beliefs: Maths** Adapted 'Early Math Beliefs and Confidence Survey' by Chen *et al.* (2014). One subscale: 'Confidence in Helping Children Aged 4–6 Learn Maths'

Due to the COVID-19 pandemic, the timeline of the efficacy trial was amended within the protocol (Ainsworth *et al.,* 2023). Initially, the efficacy trial was due to start within schools in September 2020, but this was delayed by one year until September 2021. The EEF approved the one-year delay in April 2020. At the time, a national lockdown was in place and the EEF had paused school recruitment and evaluation activities, and schools closed. Given the unknowns regarding when the lockdown would be lifted and schools would re-open, and the impact COVID-19 may have on schools' willingness to join the trial, it was not feasible to recruit schools to start the trial in September 2020. A one-year delay allowed the intended trial duration to be maintained, including the intended TEEMUP PD delivery length. School recruitment was scheduled to begin in January 2021. Lockdown restrictions for COVID-19 fully ended in July 2021, prior to the trial starting in schools in September 2021, however, other restrictions (e.g. social distancing, the wearing of face masks, and self-isolation) lasted up to 1 April 2022.

School and participant selection

Schools

The delivery team led on the recruitment of schools. School recruitment was planned to start in January 2021, however, the UK entered its third national lockdown due to the COVID-19 pandemic on 6 January 2021 and schools were closed from this point until 8 March 2021. During this time, the delivery team advertised the study on social media and through local authorities but did not contact schools directly. Direct school contact began once schools reopened in March 2021. Recruitment strategies included: emails to schools in recruitment areas; marketing through social media channels; promotion via sector press and public relations work; and working with contacts in targeted local authorities and providing them with recruitment materials to facilitate recruitment at a local level. The delivery team aimed to overrecruit schools to offset an anticipated higher than usual pre-randomisation drop-out (due to the COVID-19 pandemic), however, school recruitment was challenging. While it was planned school recruitment would end in August 2022, it continued until October 2022. Reasons for school's non-participation included: high levels of COVID-related staffing issues and illness; schools focusing on pandemic recovery; schools having signed up to other programmes/initiatives; and lack of funding for staff cover.

Copies of all recruitment documents are included in the Technical Notes accompanying this report. Initial information about the project was provided in the form of the Study Information Sheet and link to the study website as well as having direct communication with the delivery team. Schools that were interested in participating were requested to return a completed Expression of Interest form to the delivery team via email or to complete the form online. The delivery team then checked their eligibility (criteria listed below).

Once eligibility was confirmed, the delivery team emailed each school an MoU, which provided full details relating to a school's involvement within the trial, and a data sharing agreement between the school and the University of York. Schools willing to participate returned a completed and signed MoU and data sharing agreement to the delivery team who forwarded these to the evaluation team.

School inclusion criteria

- State primary and infant schools.
- Schools located in the East of England and bordering local authorities primarily, as these areas were within reasonable travel distances for the delivery team to facilitate training sessions.

- Schools that have a Reception cohort size ideally greater than 20 (excluding children where EAL and/or those with Special Educational Needs and Disabilities (SEND), which was a barrier to participation as explained further below).

- Schools willing to nominate one Reception teacher and at least one Year 1 teacher (who were full-time or majority of the time), who would participate in the TEEMUP PD if their school was randomly allocated to the intervention group.

- Schools that anticipate that nominated Reception and Year 1 staff would remain teaching the same year groups over the duration of the study.

- Schools that can commit to keeping participating children in classes taught by nominated Reception and Year 1 teachers.

- Generally, schools with more than 7% of pupils who have ever been eligible for FSM (soft recruitment target, starting with schools with more than 15% pupils ever eligible for FSM).

- Schools that agreed to all requirements outlined in the Study Information Sheet, MoU, and data sharing agreement.

School exclusion criteria

- Schools in a multi-academy trust (MAT), where another school in the MAT is taking part in TEEMUP (only one school per MAT was eligible for TEEMUP to minimise the risk of contamination). Should it have become known that a school, not previously part of a MAT, joined one part-way through the trial where that MAT includes another participating school allocated to the opposite trial arm, that school would not be withdrawn from the evaluation as this reflects a real-world possibility. The risk of contamination and level of engagement with the intervention for the control school would be assessed.

- Schools already taking part in a Reception or Year 1 substantial PD-related research study.

- Schools already taking part in a trial funded by the EEF in the early years or in Key Stage 1 (e.g. the White Rose Maths trial or the Maths Champions II trial).

- SEND schools.

- Private schools.

For clarity, schools with the following criteria/characteristics were eligible for TEEMUP:

- Schools signed up or planning to sign up to trial/pilot the 2019 new ELGs of the EYFSP.

- Larger schools (e.g. three class entry) were considered for inclusion on a case-by-case basis.

- Infant only schools.

- Participating schools (allocated to TEEMUP or control) were able to adopt other Key Stage 1 maths interventions (including PD) during the duration of the project, as part of business as usual. Details on this were collected at various time points throughout the trial via school usual practice surveys, further details are provided in the 'IPE methods' section below.

Pupils

Pupil inclusion criteria

Cohort 1

- Pupils aged four to five years, who started Reception in September 2021.

- Pupils in a class of a nominated Reception teacher.

Cohort 2

- Children aged four to five years, who started Reception in September 2022.
- Children in a class of a nominated Reception teacher.

Pupil exclusion criteria

- Pupils were not eligible to take part in the trial if teachers considered them to have significant SEND or EAL where an extreme language barrier exists or they are new to English, which would prevent them from accessing the assessment and/or cause distress through completing the assessment. Excluded pupils in participating intervention schools still experienced the intervention.

During September 2021 to October 2021, recruited schools were asked to provide the number of pupils in their school who met the eligibility criteria for Cohort 1. The evaluation team then provided each school with electronic and paper copies of parent/carer information sheets and withdrawal forms (see Technical Notes) and asked that they be distributed to the parents/carers of all eligible pupils at the school. This informed parents/carers about the trial, what it meant for their child to take part, and how to withdraw their child if they did not want them to participate. In addition to distributing paper and electronic versions of the trial parent/carer information sheets and withdrawal forms, schools were encouraged to use all their usual communication channels to ensure that information about the research was disseminated to all relevant parents/carers (e.g. letter, email, text, and school newsletter).

Parents/carers had a two-week window to withdraw their child from the evaluation elements of the project (data sharing and assessments) by completing and returning the withdrawal form to the school or contacting the school and expressing this wish verbally. Schools were advised to complete a withdrawal form on behalf of any parent/carer expressing this wish to the school through another channel (such as phone call or in-person conversation). Schools were asked to not send the evaluation team information about pupils whose parents/carers had chosen to withdraw them from the trial. Schools were advised to securely store returned withdrawal forms on-site for the duration of the project and were instructed to ensure any such children were removed from lists of participating pupils before secure transfer to the evaluation team. For participating pupils (i.e. those whose parents/carers did not chose to withdraw them from the evaluation), the school securely transferred their details to the evaluation team using the University of York Online DropOff service.

This process was repeated in Autumn Term 2022 to recruit pupils for Cohort 2.

Cohort 1 completed the BAS3 ENC twice with a research assistant, once at baseline during November 2021 to December 2021 and again for outcome assessment during June 2023 to July 2023. Cohort 2 completed the BAS3 ENC with a research assistant once in June 2023 to July 2023 for outcome assessment only. Cohort 2 did not complete the BAS3 ENC at baseline for both logistical and statistical reasons. A baseline measurement would have meant schools receiving another assessment visit in November 2022 to December 2022, increasing the burden on the schools and the trial cost. In addition, this measurement would have taken place after randomisation and therefore, would have been biased as a statistical measure of baseline maths attainment, as it could have been influenced by the allocation randomly assigned to the school. For Cohort 2, we instead aimed to adjust for a baseline measure of maths attainment through the use of a lagged-school-level measure, details of this are provided in the 'Secondary outcomes' subsection of the 'Methods' section below.

For Cohort 1, the evaluation team aimed to assess at least 15 pupils per school with the BAS3 ENC at baseline and to assess the same pupils again at post-test. For Cohort 2, the evaluation team aimed to assess 15 pupils per school for the CSBQ at baseline and post-test. The following assessment sampling process was undertaken prior to baseline assessment for Cohort 1 and prior to CSBQ baseline assessment for Cohort 2. In cases where a school had more than 15 eligible pupils within each cohort, purposive and random sampling was performed. A key priority for the funder of this evaluation, the EEF, is raising the attainment of disadvantaged pupils (i.e. those eligible for FSM). For each cohort, in order to ensure a sufficient sample size to conduct an exploratory analysis on FSM, in schools where there were three or fewer pupils eligible for FSM, all these pupils were included in the trial sample for assessment and then the remaining eligible pupils were randomly ordered for assessment. In schools with more than three pupils eligible for FSM, three pupils eligible for FSM were randomly

sampled from the FSM group for assessment, and then all remaining pupils (FSM and non-FSM) were randomly ordered for assessment.

For Cohort 1, the independent blinded research assistants who competed the BAS3 ENC were advised to work their way through the provided register (which included the first 22 pupils from the ordered list, or, if fewer than 22 pupils were eligible, all eligible pupils), starting with the child who was first on the list (up to three pupils eligible for FSM appeared first), and continue until the assessment had been completed for at least 15 pupils or more if possible on the visit date. If a pupil was absent on the day of baseline BAS3 ENC testing, the research assistant assessed the next available pupil from the ordered list. This process served to prevent unconscious assessor bias during data collection given that pupil absence was high during the baseline assessment phase, as it was a legal requirement for anyone with COVID-19 to self-isolate for ten days. When research assistants revisited the school to complete the post-test with Cohort 1, they only completed assessments with pupils who were assessed at baseline. The group of 15 Cohort 1 pupils (or more), per school, for whom baseline assessments were completed formed the randomised evaluation group. A pupil list was provided to nominated teachers to enable them to conduct the CSBQ for the relevant pupils for each cohort at each time point. At baseline, teachers were requested to complete the CSBQ for at least the first 22 pupils on the list (who were listed as essential and were the same pupils that were listed on the BAS3 ENC assessment register) and the remainder if they had capacity (NB. teachers were under significant pressure due to the COVID-19 pandemic, e.g. due to high staff absences due to COVID-19; Ofsted, 2021, and the evaluation team were keen to keep the evaluation burden to teachers to a minimum). At the follow-up time points for Cohort 1 (June 2022 to July 2022 and June 2023 to July 2023), teachers were requested to complete the CSBQ for all pupils who had completed the BAS3 ENC at baseline and for the remaining pupils if possible (the remaining pupils listed were those who just had the CSBQ completed as baseline).

For Cohort 2, the nominated Reception teacher at each school was provided with a list of 15 Cohort 2 pupils (i.e. the first 15 pupils who had been selected/ordered as detailed above) and were requested to complete the CSBQ for all 15 children on the list at baseline. During the CSBQ baseline assessment time point, if a teacher advised the evaluation team that a child had left or been withdrawn, then the next child on the ordered list was selected for assessment instead and the school was provided within an updated list. As post-test, research assistants who completed the BAS3 ENC were provided with the same list of 15 Cohort 2 pupils and were asked to assess all pupils on the list. As with the post-test for Cohort 1, if a Cohort 2 pupil on the assessment register was absent at post-test, they were not replaced by another pupil at that stage.

Outcome measures

Primary outcome

Maths attainment: The BAS3 ENC was the primary baseline and outcome measure (Elliot and Smith, 2011). BAS3 ENC is designed to measure young children's knowledge and application of concepts of number and quantity (concepts of measurement are not included) between the ages of three and seven years old. It assesses the following skills and concepts:

- rote number counting;
- counting objects;
- matching and classifying by qualitative attributes and by number;
- comparison sets by concepts (e.g. more, less);
- recognising number names and numerals;
- recognising ordinal relationships, e.g. first, second, and third;
- understanding numerical order;
- basic addition and subtraction; and
- counting by tens and recognising place value of tens and ones.

The BAS3 ENC comprises 30 items in total and has an item selection procedure to identify the items most appropriate to the individual child. Children start at a particular item depending on their age:

- three years to four years five months - Item 1;
- four years six months to six years five months – Item 4; and
- six years six months to seven years 11 months – Item 18.

After considering the available measures of maths attainment, the BAS3 ENC was determined to be the most appropriate primary baseline and outcome measure for this trial by both the evaluation team and delivery team. As the BAS3 ENC is delivered on a one-on-one basis, takes just five to ten minutes to complete, and is developed for children aged three to seven, it was therefore, suitable to be used at each data collection time point for maths attainment (i.e. with young pupils at the start of Reception [Cohort 1 only], the end of Reception [Cohort 2 only]), and end of Year 1 [Cohort 1 only]). Using the same measure for baseline and outcome assessments allowed the evaluation team to assume a higher pre- and post-correlation, which was advantageous for the sample size calculations. Of the available measures, BAS3 ENC was also considered the most appropriate measure for the trial in terms of the skills and concepts measured. The FEEL efficiency trial (Siraj *et al.,* 2018a) used the Differential Ability Scales-II (DAS-II) ENC (i.e. the USA-standardised version of the BAS ENC) as an outcome measure and reported that children attending settings that used the FEEL CPD programme (which the TEEMUP PD is based on) showed improvements in this measure relative to the control group; using the BAS3 ENC therefore, enabled more direct comparisons to be made between results from the current efficacy trial and the FEEL efficiency trial, compared to using other measures of maths attainment. However, it should be noted that the contexts for the studies, and, as a result the CPDs, were different. In the TEEMUP PD, qualified teachers were the participants and before they started the CPD, they taught maths as part of their daily routines. Teachers were expected to teach maths in both Reception and Year 1 in England (see the National Curriculum and Early Years Foundation Stage Guidance[1]). This was not the case in the early years settings in the FEEL study, maths was not an expectation in the Australian early years learning framework, and teaching maths was new to many participants.

The BAS3 ENC is scored between 0 and 35. This raw score is then converted to a T-score (maximum range 20–80) based on item starting point and age using the GL Assessment pre-defined process. A higher score indicates greater attainment.

The assessment was administered by an independent, blinded, trained research assistant. The assessment predominately requires the research assistant to ask the child questions about pictures presented in a stimulus booklet. The child answers questions by pointing, counting aloud, or saying a number. None of the answers require extended verbal responses from the child. The assessment takes five to ten minutes per pupil to complete. As well as indicating knowledge of the skills and concepts listed above, higher BAS3 ENC scores may also be indicative of a child's verbal comprehension, basic language concepts, visual perception and analysis of pictures, and integration of visuals and verbal conceptual information. It should be noted that a low BAS3 ENC score may reflect a poor knowledge of the skills and concepts listed above, and/or expressive language difficulties, including a reluctance to speak (Elliot and Smith, 2011).

The research assistants received training from an educational psychologist and the evaluation team on how to deliver the assessment. All research assistants had an enhanced Disclosure and Barring Service check and underwent relevant safeguarding and data protection training. Assessment results were recorded 'live' electronically via Qualtrics survey software, with paper collection available as a backup. The evaluation team advised that a familiar staff member be available to chaperone the assessment conducted by the research assistant to ensure the pupils were comfortable. If a research assistant became unblinded (e.g. the school mentioned that they had been following the TEEMUP PD programme), they were asked to email the evaluation team so this could be recorded (but there were no known instances of this).

[1] https://www.gov.uk/government/publications/early-years-foundation-stage-framework--2

Secondary outcomes

CSBQ

The 34-item CSBQ, developed by Howard and Melhuish (2017), was completed by teachers to collect data on children's: i) self-regulation, that is a child's ability to control their own thoughts, behaviour, reactions, and interactions; and ii) PSED overall. The CSBQ is designed to be completed by a parent or educator who knows the child, so it was appropriate to ask teachers to complete this measure. The CSBQ was chosen as an outcome measure as it is quick to complete by teachers, taking around five minutes per child, and provided data to address the research questions relating to children's self-regulation and PSED. The CSBQ was also used as an outcome measure in the FEEL efficiency trial (Siraj *et al.,* 2018a) discussed in the section above. The CSBQ pertains to children's everyday behaviours related to social and emotional development and self-regulation. It yields seven subscales that all contain at least five items (some items are used in more than one subscale):

1. Cognitive self-regulation (items 5, 6, 8, 12, 18).
2. Emotional self-regulation (items 2, 10, 11, 14, 23, 26).
3. Behavioural self-regulation (items 7, 13, 15, 29, 30, 31).
4. Sociability (items 1, 4, 9, 16, 22, 27, 32).
5. Prosocial behaviour, for example, sharing, showing empathy (items 15, 19, 24, 27, 30).
6. Externalising problems, for example, being antisocial (items 3, 20, 23, 26, 28).
7. Internalising problems, for example, being anxious or depressed (items 17, 21, 25, 33, 34).

For each item, the respondent is asked to evaluate the child's frequency of target behaviours on a five-point scale (1 = 'not true' to 3 = 'partly true' to 5 = 'very true'). The items are either positively worded, for example, 'Persists with difficult tasks' and 'Chooses activities on their own', or negatively worded, for example, 'Regularly unable to sustain attention' and 'Not able to sit still when necessary'. The items are scored whereby the higher the child scores on these scales, the more they show these behaviours. For subscales 1 to 5, scores on negatively worded items are reversed prior to analysis. For subscales 6 to 7, negatively worded items are not reversed before analysis, therefore, the higher the children score on these scales, the more they show externalising and internalising problems. The seven subscale scores are obtained by taking the average of the component item scores (first reversing any relevant score). The developers offer no guidance on how to handle missing item-level data for this instrument and so the subscales were only scored if a valid response was provided to all items. Cronbach's alphas for each subscale have previously been reported as follows: sociability = 0.74; internalising = 0.78; emotional self-regulation = 0.83; cognitive self-regulation = 0.87; externalising = 0.88, prosocial = 0.89; and behavioural self-regulation = 0.89. The CSBQ subscales have internal consistency ranging from 0.74 to 0.89, and concurrent validity in the range of 0.48 to 0.81 (Howard and Melhuish, 2017).

To assess self-regulation, we used a single overall index of children's self-regulatory capacities that represents the mean of the CSBQ's three subscales of cognitive self-regulation, emotional self-regulation, and behavioural self-regulation.

To assess PSED, we considered each of the seven CSBQ subscales separately.

For Cohort 1, teachers were asked to complete the CSBQ at the start and end of Reception (November 2021 to December 2021 and June 2022 to July 2022) and again at the end of Year 1 (June 2023 to July 2023). We also asked the participating Reception teacher to complete the CSBQ at the start and end of Reception for Cohort 2 (November 2022 to January 2023 and June 2023 to July 2023). At each time point, the evaluation team securely provided teachers with a list of the participating pupils and their assessment IDs for whom we were requesting the CSBQ to be completed for. Teachers were requested to complete the CSBQ electronically via Qualtrics.

EYFSP

The EYFSP is an observational measure completed by teachers when children are in the Summer Term of Reception (DfE, 2023b). Therefore, this was assessed after five to six months of intervention for Cohort 1 pupils. Access to relevant EYFSP data was requested from the NPD, via the ONS SRS. Schools and parents/carers were made aware that their data would be

'matched' to NPD data in the MoU and parent/carer information sheet. The EYFSP measures 17 ELGs whereby the teacher assigns the child as being 'emerging' or 'expected' for each (DfE, 2023b).

Mathematics ELGs

Number: 'Children at the expected level of development will: have a deep understanding of number to 10, including the composition of each number; subitise (recognise quantities without counting) up to 5; and automatically recall (without reference to rhymes, counting or other aids) number bonds up to 10, including double facts' (DfE, 2024, p. 26).

Numerical Patterns: 'Children at the expected level of development will: verbally count beyond 20, recognising the pattern of the counting system; compare quantities up to 10 in different contexts, recognising when one quantity is greater than, less than or the same as the other quantity; and explore and represent patterns within numbers up to 10, including evens and odds, double facts and how quantities can be distributed equally' (DfE, 2024, pp. 26–27).

ELG Number and Numerical Patterns were combined and analysed as a binary outcome (expected level met for both ELGs).

Self-regulation ELG

Self-regulation: 'Children at the expected level of development will: show an understanding of their own feelings and those of others, and begin to regulate their behaviour accordingly; set and work towards simple goals, being able to wait for what they want and control their immediate impulses when appropriate; and give focused attention to what the teacher says, responding appropriately even when engaged in activity, and show an ability to follow instructions involving several ideas or actions' (DfE, 2024, p. 24).

This outcome was analysed as a binary variable (expected level met).

PSED ELGs

Self-regulation: As above.

Managing self: 'Children at the expected level of development will: be confident to try new activities and show independence, resilience and perseverance in the face of challenge; explain the reasons for rules, know right from wrong and try to behave accordingly; and manage their own basic hygiene and personal needs, including dressing, going to the toilet and understanding the importance of healthy food choices' (DfE, 2024, p. 25).

Building relationships: 'Children at the expected level of development will: work and play cooperatively and take turns with others; form positive attachments to adults and friendships with peers; and show sensitivity to their own and to others' needs' (DfE 2024, p. 25).

ELGs self-regulation, Managing self, and Building relationships were combined and analysed as a binary outcome (expected level met for all three ELGs).

General attainment/development

Additionally, the current EYFSP provides a general measure of good development defined as achieving: 'the expected level for the ELGs in the prime areas of learning (which are: communication and language; personal, social and emotional development; and physical development) and the specific areas of mathematics and literacy' (DfE 2024, p. 6).

This outcome was analysed as a binary variable.

We also considered the average total point score for the 17 ELGs, each assigned a score of 1 for Emerging and 2 for Expected.

Teacher confidence: Maths

Teacher confidence in teaching children maths was assessed, for nominated Reception and Year 1 teachers, using an adapted short survey 'Early Math Beliefs and Confidence Survey' by Chen *et al.* (2014). Increasing teachers' confidence in teaching maths is a key focus of the TEEMUP PD. The survey was completed at baseline during September 2021 to October 2021, and again during June 2022 to July 2022 (after five to six months of intervention), and during June 20223 to July 2023 (after the full 17 months of intervention). To try to ensure the surveys were distributed to the appropriate nominated teachers at each time point, the evaluation team updated their records of nominated teachers throughout the trial period, using information on teacher changes received either from schools directly or from the delivery team. When the teacher confidence survey was distributed during June 2022 to July 2022 and June 2023 to July 2023, schools were asked to distribute the survey to the relevant teachers if those we listed and included in our email with the survey link had since changed. The original survey consists of three subscales: 'Beliefs about Children Aged 4–6 and Maths' (five items); 'Confidence in Helping Children Aged 4–6 Learn Maths' (11 items); and 'Confidence in Own Maths Abilities' (nine items). However, only the second subscale, 'Confidence in Helping Children Aged 4–6 Learn Maths', was administered. Teachers were asked to rate their agreement with each item on a Likert scale, from 1=strongly disagree to 5=strongly agree. Scores for items in the subscale are summed to produce a summary score ranging from 11 to 55, and a higher score indicates greater confidence. The developers offer no guidance on how to handle missing item-level data for this instrument and so the scale was only scored if a valid response was provided to all 11 items.

Sample size

Sample size calculations were conducted in Stata version 17 (StataCorp LLC, College Station, TX, USA).

From protocol

For the primary analysis, we compared BAS3 ENC scores at the end of Year 1 for Cohort 1 between the intervention and control groups, adjusted for baseline BAS3 ENC score measured when the pupils were at the start of Reception. For the sample size calculation, we made the following assumptions: a school-level intracluster correlation coefficient (ICC) of 0.15, 15 pupils per school (at randomisation), a baseline and outcome test correlation of 0.6 and 1:1 allocation at the school level. The ICC and test correlation can be justified based on the following previous trials funded by the EEF, though these do differ slightly in the age of the population to this trial. The 1stClass@Number evaluation in Year 2 pupils found an ICC of 0.22 for its primary outcome of the Quantitative Reasoning Test (Nunes *et al.,* 2018) with a pre- and post-correlation of 0.29 (sample restricted to those struggling with maths) but 0.63 for the whole sample, which better reflects our population; for the secondary outcome of Key Stage 1 maths, the ICC was 0.15 with a pre- and post-correlation of 0.26 for the restricted sample but 0.63 for the whole sample. The Mathematical Reasoning evaluation in Year 2 pupils found an ICC of 0.11 for its primary outcome of the GL Assessment Progress Test in Maths with a pre- and post-correlation of 0.58 (Stokes *et al.,* 2018). The Maths Champions trial in early years, in the year before pupils started primary school, found an ICC of 0.17 for its primary outcome of the Centre of Evaluation and Monitoring (CEM) ASPECTS assessment in maths with a pre- and post-correlation of 0.59 (Robinson-Smith *et al.,* 2018).

Based on 100 schools (i.e. 1,500 pupils), we would have 80% power to show an effect size of 0.21 of a standard deviation (SD) between the control and the intervention groups in the primary analysis, allowing for 15% attrition at pupil level at post-test (Table 11). The trial aimed to recruit 106 schools to allow for some school-level attrition without comprising the minimum detectable effect size (MDES).

Based on the sampling strategy, we might conservatively assume that we will achieve an average of three EVER6FSM pupils per school (300 from 100 schools). Assuming a baseline and outcome test correlation of 0.6, an ICC of 0.15, and 15% attrition at the pupil level, with 100 schools we would have 80% power to show an effect size of 0.31 in the EVER6FSM subgroup for Cohort 1.

At randomisation

The 'randomised' sample is defined as pupils from Cohort 1 who completed the baseline BAS3 ENC. In total, 93 settings were randomised, from which there were 1,566 pupils for whom valid baseline BAS3 ENC data were available (intervention, *n*=780; control, *n*=786). This is an average of 17 pupils per school. Assuming a correlation of 0.6 between baseline and

outcome, an ICC of 0.15, and 15% pupil-level attrition, the MDES with this sample size would be approximately 0.22 (Table 11).

A total of 303 of the randomised pupils were eligible for EVER6FSM (average of 3.7 per school). With this sample size, under the same assumptions, the MDES would be approximately 0.32.

Randomisation

Schools were randomised after child recruitment and baseline data collection had been completed in that school. A statistician at the York Trials Unit randomised schools 1:1 to either the intervention arm (offered the TEEMUP PD programme) or the control arm (continued with usual provision for the duration of the evaluation). A dedicated computer program, MinimPy (Saghaei and Saghaei, 2011), was used for randomisation via minimisation using the factors:

- **School geographic location – six levels:** Peterborough, Norwich, Newmarket/Bury St Edmunds, Milton Keynes, Oxford, and Barnet, for logistical reasons, to ensure a balanced spread of intervention and control schools in each area.

- **School deprivation level:** The percentage of pupils eligible for FSM (EVER6FSM) in the school (latest available data; dichotomised at the median for the 95 schools that expressed interest in the trial, ≤16%, >16%) to ensure balance between the randomised groups, since this school characteristic and individual child deprivation may moderate outcomes.

- **School EAL level:** The percentage of pupils identified as having EAL in the school (latest available data; dichotomised at the median for the 95 schools that expressed interest in the trial, ≤8%, >8%) to ensure balance between the randomised groups, since this school characteristic and individual child EAL status may moderate outcomes.

Randomisation was carried out in batches (groups of schools that were ready to be randomised at that time) to avoid delays in programme induction and to maximise programme delivery for as many schools as possible. This was particularly necessary here as recruitment was due to end in August 2021 but was extended due to challenges in recruitment caused by COVID-19 to October 2021. Naïve minimisation with base probability 1.0 was conducted (i.e. 1:1 deterministic minimisation). Naïve minimisation was deemed to be sufficient as the allocations were conducted in batches, rather than one-by-one prospectively, meaning predictability was not a concern and hence, a random element was not required. In total, 93 settings were randomised (intervention, *n*=47; control, *n*=46) in five 'batches' during November 2021 to December 2021.

Statistical analysis

Analysis followed the EEF's (2022a) most recent guidance, and is detailed in full in the published Statistical Analysis Plan (Baird and Fairhurst, 2023). Analyses were conducted in Stata v17. The trial statistician was not blind to group allocation. All analyses were conducted on an intention-to-treat (ITT) basis, where data were available, including all schools and pupils in the group to which they were randomised irrespective of whether or not they actually received the intervention, using two-sided tests at the 5% significance level. A CONSORT diagram shows the flow of schools and pupils through the trial.

School and pupil characteristics and outcome measures assessed at baseline are summarised descriptively by randomised group both as randomised and as analysed in the primary analysis. No formal comparison of the baseline data was undertaken, except for a comparison of the difference in prior attainment (BAS3 ENC and CSBQ scores) between the groups, reported as a Hedges' g effect size with a 95% confidence interval (CI).

All outcome data are summarised descriptively by trial arm for each assessment point. The correlation of outcome measures and measures of prior attainment are presented with a 95% CI. Effect sizes based on the difference between the groups at the outcome assessment point are presented as adjusted mean differences for continuous outcomes, and odds ratios (ORs) and difference in proportions for dichotomous outcomes, with their associated 95% CI and p-value. Treatment effects are also presented as (estimated) Hedges' g effect sizes.

Primary analysis

Cohort 1 BAS3 ENC score was analysed using a mixed effects linear regression model at the pupil level. Group allocation, baseline BAS3 ENC score, and the minimisation factors (geographical location of school, EVER6FSM, and EAL) were included as fixed effects in the model, and school as a random effect. Robust standard errors were specified to account for any potential heteroscedasticity. While FSM and EAL were used as aggregate measures at the school level in the minimisation, pupil-level indicators of EVER6FSM and EAL were included in the analysis model, since these are more granular measures and so are likely to correlate better with the outcome than the school-level data.

Model assumptions were checked as follows: the normality of the standardised residuals was checked using a quantile-quantile (QQ) plot. If the model assumptions were in doubt, sensitivity analyses were conducted in which transformations of the outcome and/or covariate data were tried to improve the model fit.

Secondary analysis

Cohort 1

CSBQ

CSBQ scores were analysed via a mixed effects linear regression model incorporating both outcome time points for each child, adjusting for respective baseline CSBQ score, location, EVER6FSM, and EAL indicators, group allocation, time, and group by time interaction as fixed effects, and school and child as random effects to account for the repeated measures over time (using an unstructured covariance structure).

EYFSP

The EYFSP dichotomous measures (Mathematics ELGs, Self-regulation ELG, PSED ELGs, and Good Level of Development) were compared using mixed effects logistic regression at the pupil level, adjusted for group allocation, baseline BAS3 ENC score, location, EVER6FSM, and EAL as fixed effects, and school as a random effect. The treatment effect expressed as an adjusted OR is reported with a 95% CI and p-value, in addition to the unadjusted and adjusted percentage point difference between the two groups with a 95% CI (Ge, *et al.*, 2011). The adjusted OR (and 95% CI limits) is converted to an estimated Hedges' g effect size using the Cox index as follows (What Works Clearinghouse[2]):

$$d_{Cox} = \omega[ln(OR)]/1.65$$

Where $\omega = \left[1 - {}^{3}/_{(4N-9)}\right]$ and *N* is the total sample size.

The continuous measure (average total point score for the 17 ELGs) was analysed as described for the BAS3 ENC primary outcome.

Cohort 2

BAS3 ENC

Individual baseline BAS3 ENC scores were not collected for Cohort 2, as this would have had to have taken place post-randomisation and would also have increased the burden on schools. Therefore, we considered the use of a lagged school-level measure of prior attainment for these pupils as follows. We calculated the mean baseline BAS3 ENC score per school from Cohort 1 and the correlation between this and the outcome for Cohort 2.

Maths attainment for pupils in the intervention group and those in the control group was compared using a mixed effects linear regression model at the pupil level. Group allocation, baseline BAS3 ENC score (school-level mean from previous

[2] https://ies.ed.gov/ncee/wwc/docs/referenceresources/wwc_procedures_handbook_v4.pdf

year group), location, EVER6FSM, and EAL were included as fixed effects in the model, and school as a random effect. We conducted analyses with and without the BAS3 ENC measure as a school-level covariate in the analysis for Cohort 2.

CSBQ

CSBQ scores were analysed via a mixed effects linear regression model adjusting for respective baseline CSBQ score, location, EVER6FSM, and EAL as fixed effects, and school as a random effect.

EYFSP

The EYFSP dichotomous measures (Mathematics ELGs, Self-regulation ELG, PSED ELGs, and Good Level of Development) were compared using mixed effects logistic regression at the pupil level, adjusted for group allocation, baseline BAS3 ENC score (school-level mean from previous year group), location, EVER6FSM, and EAL as fixed effects, and school as a random effect. The treatment effect expressed as an adjusted OR is reported with a 95% CI and p-value, in addition to the unadjusted and adjusted percentage point difference between the two groups with a 95% CI (Ge, *et al.,* 2011). The adjusted OR (and 95% CI limits) is converted to an estimated Hedges' g effect size using the Cox index as follows (What Works Clearinghouse):

$$d_{Cox} = \omega[ln(OR)]/1.65$$

Where $\omega = \left[1 - {}^{3}/_{(4N-9)}\right]$ and N is the total sample size.

The continuous measure (average total point score for the 17 ELGs) was analysed as described for the Cohort 2 BAS3 ENC outcome.

These analyses were repeated omitting the BAS3 ENC covariate.

Teachers

Responses to items in the teacher confidence survey are summarised descriptively by trial arm. The summary score was compared between the two arms using mixed effects linear regression, incorporating the midpoint and endpoint scores as outcomes and adjusting for baseline score, the school-level minimisation factors (location, %FSM, %EAL) and number of years' experience as a teacher as fixed effects, and school and teacher as random effects.

Estimation of effect sizes

Effect sizes for continuous outcomes were calculated by dividing the adjusted mean difference between the intervention and control group (accounting for baseline measures and the minimisation factors) by the pooled unconditional SD obtained from the model run without these covariates. The 95% CI for the effect size was calculated by dividing the 95% confidence limits for the adjusted mean difference by this same denominator. All parameters used in these calculations are provided in this report.

$$\mathrm{ES} = \frac{(\overline{\mathrm{Y}}_{\mathrm{T}} - \overline{\mathrm{Y}}_{\mathrm{C}})_{\mathrm{adjusted}}}{sd_{pooled}}$$

where, $(\overline{\mathrm{Y}}_{\mathrm{T}} - \overline{\mathrm{Y}}_{\mathrm{C}})_{\mathrm{adjusted}}$ denotes the difference in means between trial groups adjusting for pre-test score and the minimisation factors, from the multilevel analysis model; and sd_{pooled} denotes the pooled, unconditional SD of the two groups (square root of the sum of the within- and between-cluster variances).

Estimation of ICCs

The ICC associated with school for the outcomes (both pre- and post-test where available) are presented alongside a 95% CI. The ICC at post-test was computed for the analysis model, and also for an empty model (i.e. one without covariates). The ICC at pre-test was calculated for a linear model with pre-test as the outcome and setting as a random effect.

Subgroup analyses

For both cohorts, subgroup analyses looking at gender, EVER6FSM eligibility, and whether schools were taking part in the National Centre for Excellence in the Teaching of Mathematics (NCETM) Maths Hubs Programme were undertaken for the BAS3 ENC outcome.
In surveys conducted at baseline (Autumn Term 2021), midpoint (Autumn Term 2022), and end of study (Summer Term 2023), the TEEMUP lead contact within each school was asked about their school's participation in the NCETM Maths Hubs Programme (schools can join one of 40 Maths Hubs, which run face-to-face CPD and online support). At baseline, the senior management team (SMT) staff were asked if their school had signed up to NCETM Maths Hub in the 2021–2022 academic year; at midpoint, lead contacts were asked if schools were signed up to the NCETM Maths Hubs Programme in the 2022–2023 academic year; and at endpoint, lead contacts were asked if schools had signed up to the NCETM Maths Hubs Programme in the 2022–2023 academic year. For the purposes of analyses, a school was considered to be taking part in the NCETM Maths Hubs Programme if they completed at least one of these surveys and responded 'Yes' to these questions at any time point.

The subgroup analyses were conducted by including the factor and an interaction term between the factor and allocation in the primary analysis model. We also repeated the primary analysis within the subset of participants eligible for EVER6FSM.

Analysis in the presence of non-compliance

CACE analysis was conducted on the primary outcome of maths attainment as measured by the BAS3 ENC on Cohort 1 and Cohort 2. These CACE analyses aimed to obtain a treatment effect estimate among 'compliers', which may differ from the primary ITT analysis. This particular trial does not differentiate between compliance and fidelity for the CACE analysis and seeks to capture information on both compliance and fidelity within one measure. The IPE sought to explore compliance and fidelity as separate constructs where possible.

As suggested by the EEF analysis guidance (2022a), two thresholds for compliance are defined to conduct two CACE analyses for 'good' compliance and at least 'minimal' compliance.

As described in the 'Changes to the intervention due to the COVID-19 pandemic' the TEEMUP PD was altered during the trial to make it more accessible to the teachers. As such, the compliance criteria were also diluted (as detailed in the published protocol, Ainsworth *et al.,* 2023) to reflect the challenges of staff retention and the demands on teachers' time:

- Rather than requiring that children were taught mathematics over 100% of the intervention by a TEEMUP trained teacher, this was reduced to 50%.
- Teachers could self-report having watched the online recording of a session as a substitute to attending a face-to-face or online session.
- Rather than requiring that all children in Cohort 1 moved from a TEEMUP trained Reception teacher to a TEEMUP trained Year 1 teacher, this was reduced to 75% or at least 15 pupils, whichever was lower.

Cohort 1

Compliance was measured at the school level since the impact of the intervention depended on engagement of both the Reception and Year 1 teachers that the pupils in Cohort 1 will have been taught by. Each teacher in the intervention arm was assessed for their compliance with the intervention. A school was classed as having 'good' compliance if they fulfilled all the core criteria in Table 5.

Table 5: Cohort 1 CACE analysis GOOD compliance 'core' criteria

GOOD compliance 'core' criteria	Data collection by/from
TEEMUP trained Reception and Year 1 teachers complete seven of the first nine core training workshops, at least by watching recorded sessions. (NB. Attendance/watching final half-day session is not required for compliance), and the TEEMUP trained Reception teacher/s remains at the school and teaching Reception during the majority (>50%) of the 2021–2022 academic year and TEEMUP trained Year 1 teacher/s remains at the school and teaching Year 1 for the majority (>50%) of the 2022–2023 academic year	Attendance at training collected by the delivery team via attendance registers/training completion records for each school and shared with the evaluation team Teachers and teacher changes collected by the evaluation team directly from schools at the end of each academic year, and by the delivery team through PD. Lists shared and cross-referenced between the two teams
The school hosts three face-to-face visits from a mentor/coach, at least two of which teachers should be well prepared for (the delivery team will define 'preparedness', which will consist of two elements, ensuring there is time during the meeting to: i) review existing change plans and write/agree new ones; and ii) gather evidence of changes made relating to previous agreed actions and/or TEEMUP PD	Collected by the delivery team mentor records for each school and shared with the evaluation team
A minimum of eight school logins to the online knowledge base over the course of the whole intervention period	Automated data held by the delivery team shared with the evaluation team or self-reported data from teachers collected by the delivery team and shared with the evaluation team
>75% or at least 15 pupils, whichever is lower (e.g. 12 or more of 15, and 15 or more of 22) of pupils in the evaluation (Cohort 1 Reception pupils for whom a baseline BAS3 ENC assessment was conducted) move to a Year 1 class being taught maths by a TEEMUP trained Year 1 teacher in the 2022–2023 academic year	Evaluation team
School's TEEMUP mentor considers the school to have been 'good' compliers, i.e. the school can provide sufficient evidence of change in practice resulting from TEEMUP training/resources. The mentors will assess evidence of change in practice on a scale of 0 to 3 with 0=no change, 1=minimal change, 2=good change, and 3=excellent change. A school must score at least 2 to be classed as a good complier	Collected by the delivery team mentor records for each school and shared with the evaluation team

A school was classed as having at least 'minimal' compliance if they fulfilled all the criteria detailed in Table 6.

Table 6: Cohort 1 CACE analysis MINIMAL compliance criteria

MINIMAL compliance criteria	Data collection by/from
TEEMUP trained Reception and Year 1 teachers complete five of the first nine core training workshops, at least by watching recorded sessions. (NB. attendance/watching final half-day session is not required for compliance) and a TEEMUP trained Year 1 teacher/s remains at the school and teaching Year 1 for the majority (>50%) of the 2022–2023 academic year	Attendance at training collected by the delivery team via attendance registers/training completion records for each school and shared with the evaluation team Teachers and teacher changes collected by the evaluation team directly from schools at the end of each academic year, and by the delivery team through PD. Lists shared and cross-referenced between the two teams
The school hosts two face-to-face visits from a mentor/coach, at least one of which teachers should be well prepared for (the delivery team will define 'preparedness' which may include: class cover arranged; an appropriate meeting place organised; read through questions provided by the delivery team prior to meeting; and prepared to answer them)	Collected by delivery team mentor records for each school and shared with the evaluation team
A minimum of four school logins to the online knowledge base over the course of the whole intervention period	Automated data held by the delivery team shared with the evaluation team or self-reported data from teachers collected by the delivery team and shared with the evaluation team
>50% or at least 11 pupils, whichever is lower (e.g. 8 of 15 pupils, or 11 of 22) of pupils in the evaluation (Cohort 1 Reception pupils for whom a baseline BAS3 ENC assessment was conducted) move to a Year 1 class being taught maths by a TEEMUP trained Year 1 teacher in the 2022–2023 academic year	Evaluation team

Cohort 2

Compliance was measured primarily at the teacher level since the impact of the intervention depended on engagement of the Reception teacher only for Cohort 2. Each Reception teacher in the intervention arm was assessed for their compliance with the intervention.

A Reception teacher was classed as having 'good' compliance if they fulfilled all of the core criteria in Table 7.

Table 7: Cohort 2 CACE analysis GOOD compliance 'core' criteria

GOOD compliance criteria 'core' criteria	Data collection by/from
TEEMUP trained Reception teacher/s completes seven of the first nine core training workshops, at least by watching recorded sessions. (NB. attendance/watching final half-day session is not required for compliance), and the TEEMUP trained Reception teacher/s remains at the school and teaching Reception for the majority (>50%) of the 2022–2023 academic year	Attendance at training collected by the delivery team via attendance registers/training completion records for each school and shared with the evaluation team Teachers and teacher changes collected by the evaluation team directly from schools at the end of each academic year, and by the delivery team through PD. Lists shared and cross-referenced between the two teams
The school hosts three face-to-face visits from a mentor/coach, at least two of which the Reception teacher should be well prepared for (delivery team will define 'preparedness', which may include class cover arranged, an appropriate meeting place organised, read through questions provided by the delivery team prior to meeting and prepared to answer them). It is acceptable for these visits to be conducted with different teachers (if there is a change in Reception teacher in the school between 2021–2022 and 2022–2023), provided at least two are conducted with a TEEMUP trained teacher	Collected by delivery team mentor records for each school and shared with the evaluation team
A minimum of eight school logins to the online knowledge base over the course of the whole intervention period	Automated data held by the delivery team shared with the evaluation team or self-reported data from teachers collected by the delivery team and shared with the evaluation team
School's TEEMUP mentor considers the school to have been 'good' compliers, i.e. the school can provide sufficient evidence of change in practice resulting from TEEMUP training/resources. The mentors will assess evidence of change in practice on a scale of 0 to 3 with 0=no change, 1=minimal change, 2=good change, and 3=excellent change. A school must score at least 2 to be classed as a good complier	Collected by delivery team mentor records for each school and shared with the evaluation

A Reception teacher was classed as having at least 'minimal' compliance if they fulfilled all of the criteria as detailed in Table 8.

Table 8: Cohort 2 CACE analysis MINIMAL compliance criteria

MINIMAL compliance criteria	Data collection by/from
TEEMUP trained Reception teacher/s completes five of the first nine core training workshops, at least by watching recorded sessions. (NB. attendance/watching final half-day session is not required for compliance) and the TEEMUP trained Reception teacher/s remain at the school and teaching Reception for the majority (>50%) of the 2022–2023 academic year	Attendance at training collected by the delivery team via attendance registers/training completion records for each school and shared with the evaluation team Teachers and teacher changes collected by the evaluation team directly from schools at the end of each academic year, and by the delivery team through PD. Lists shared and cross-referenced between the two teams
The school hosts two face-to-face visits from a mentor/coach, at least one of which Reception teacher should be well prepared for (delivery team will define 'preparedness', which may include class cover arranged, an appropriate meeting place organised, read through questions provided by the delivery team prior to meeting, and prepared to answer them). It is acceptable for these visits to be conducted with different teachers (if there is a change in Reception teacher in the school between 2021–2022 and 2022–2023) as long as at least one of these mentor meetings is with the trained Reception teacher	Collected by delivery team mentor records for each school and shared with the evaluation team

A minimum of four school logins to the online knowledge base over the course of the whole intervention period	Automated data held by the delivery team shared with the evaluation team or self-reported data from teachers collected by the delivery team and shared with the evaluation team

Two CACE analyses (Dunn, Maracy, and Tomenson, 2005) for each cohort were conducted for the BAS3 ENC outcome defining compliance of the schools as a dichotomous variable in the two ways described above. These analyses used a Two Stage Least Squares (2SLS) approach with group allocation as the instrumental variable for the compliance indicator, with cluster standard errors to account for clustering at the school level. CACE analyses were conducted at the pupil level. Results for the first stage (which predicts the compliance indicator using the treatment allocation as instrumental variable alongside all other covariates included in the second stage) are reported alongside: i) the correlation between the instrument and the endogenous variable; and ii) an F-test.

Missing data analysis

The amount of missing baseline and outcome data is summarised, and reasons for missing data explored and provided in the report where available. Since the percentage of missing cases exceeded 5%, a multilevel logistic regression model was used to model presence or absence of the primary outcome including all available pupil- and school-level baseline data as fixed effects, and school as a random effect. Significant predictors are discussed.

The impact of missing data on the primary analysis was additionally assessed using multiple imputation (MI) including all available pupil- and school-level baseline variables (school: location, percentage of pupils every eligible for EVER6FSM, percentage of pupils with EAL; pupil: gender, EVER6FSM status, EAL status, BAS3 ENC, and CSBQ scores).

A 'burn-in' of ten was used, which means that the first ten iterations of the imputation were not used to allow the iterations to converge to a stationary distribution, and 30 imputed datasets were created. The primary analyses was then rerun within the imputed datasets and Rubin's rules (Rubin, 1987) used to combine the multiple imputed estimates.

Additional analyses and robustness checks

In sensitivity analyses, the analysis models for EYFSP Self-regulation and PSED outcomes were repeated adjusting for baseline CSBQ scores for these domains, rather than baseline BAS3 ENC score, since they propose to measure the same domains and so we anticipated a reasonable correlation between baseline score and outcome. The correlations were calculated and reported.

A further sensitivity analysis was included to adjust for whether or not the participating school is involved in the NCETM Maths Hubs Programme for the BAS3 ENC outcome, in both cohorts.

Baseline CSBQ scores for Cohort 2 were collected between November 2022 and January 2023. Children's abilities may have changed over this time. Therefore, we considered the proportion of baseline CSBQ data that were collected before and after Christmas, and present this by trial arm. In a sensitivity analysis for the Cohort 2 CSBQ analysis, we included an indicator for whether the data were collected before or after Christmas.

Implementation and process evaluation

The IPE followed the EEF principles and guidance for undertaking process evaluations (Humphrey *et al.,* 2016; EEF, 2022b). The primary aim was to evaluate fidelity of implementation, facilitators, and/or barriers to implementation, and to gather key stakeholders' perceptions of the intervention. The IPE aimed to explore the programme theory and relationships between different components of the logic model with an aim to test its causal assumptions and inform intervention scale-up. This IPE was designed to complement the impact evaluation by *explaining* any observed effect of the TEEMUP PD intervention on children's maths outcomes.

Research methods

Surveys

- All participating schools were requested to complete surveys to capture usual practice at baseline (Autumn Term 2021), mid-implementation (Autumn Term 2022), and at the end of the trial (Summer Term 2023). The surveys established usual practice over the duration of the trial to determine whether any changes to practice were made in control schools, which may influence maths teaching and learning in Reception and Year 1. Where appropriate, surveys also collected data from nominated Reception and Year 1 teachers in intervention schools on implementation and costs associated with delivering the TEEMUP PD.

Interviews

Interviews took place at various points throughout the two-year trial and involved semi-structured interviews with key school-based stakeholders, for example, nominated Reception and Year 1 teachers, senior management, or other relevant school staff. Schools were recruited to participate in one of two types of case study interviews:

- **Longitudinal**: The purpose of the longitudinal case studies was to follow schools for the entirety of the TEEMUP PD journey. Intended interview time points (Summer Term 2022, Autumn Term 2022, Spring Term 2023, and Summer Term 2023) offered the opportunity for stakeholders to regularly reflect and provide feedback on the delivery of the TEEMUP PD, to chronologically capture any barriers, facilitators, and adaptations during implementation, and report on changes to usual practice, perceived impact, and suggested improvements to the intervention. It was also intended to observe one mentor session within longitudinal case study schools, with school staff providing separate consent for this element of the IPE. Intervention schools were purposively selected to be invited to interview based on the percentage of pupils within the school in receipt of FSM with the aim of having a balance of schools, which had low-medium rates of FSM (≤18%) with schools that had high rates of FSM (above 18%).

- **By compliance:** The purpose of these case studies was to help unpick what the necessary conditions (enabling factors) are for participating teachers and schools to engage with and implement the intervention as intended and understand how compliance may impact on outcomes. Schools were identified by the delivery team using the compliance data available at the end of academic year 2021–2022 to participate in interviews during Autumn Term 2022 and Summer Term 2023. Using the CACE criteria, schools which were identified as having 'good' compliance with the TEEMUP PD, to provide examples of 'best practice', were invited to interview. Similarly, invites were sent from the pool of schools where the CACE criteria indicated they had 'minimal' compliance with the TEEMUP PD.

All schools were invited to email to participate in semi-structured interviews held over the online video conferencing software, Zoom. The invitation included a participant information sheet and link to complete an eConsent form (using Qualtrics survey software). A pre-agreed semi-structured interview schedule was implemented to ensure the same questions were asked to all interview participants at the relevant time point while allowing the opportunity for the interviewers to probe interesting lines of enquiry. Interviews lasted 20–30 minutes and were audio recorded for transcription purposes only. Interviews were undertaken from May 2022 to July 2023. Copies of information sheets and consent forms used for interviews are provided in the Technical Notes.

Observations of TEEMUP PD training workshops

Observations of the face-to-face TEEMUP PD training workshops were conducted to allow the evaluation team to gain a better understanding of the intervention. They also offered the opportunity to gain insight into how the day ran and to have light-touch conversations with: i) participating Reception and Year 1 teachers and senior management to gauge perspective of engagement, responsiveness, and quality of the PD, and inform fidelity; and ii) the delivery team to gain their feedback and reflections on delivery methods and materials. The evaluation team planned to observe the following TEEMUP PD training:

- Full-day workshops 1 and 2 (scheduled for Spring Term 2022).

- Half-day workshop 9 (scheduled for Spring Term 2022).

- Half-day workshop 10 (scheduled for Summer Term 2023). This observation was not undertaken due to staff illness within the evaluation team.

Delivery team

In-depth semi-structured interviews with the delivery team were conducted at two time points: i) Summer Term 2022 (mid-implementation); and ii) Summer Term 2023 (post-implementation). Informed consent was sought from all participants prior to the interviews.

Monitoring data

The following monitoring data was collected by the delivery team through administrative records and in-school mentor visits and shared with the evaluation team:

- completion of teacher training workshop (whether face-to-face, online catch-up, or absent);
- frequency of 'core' and ad hoc mentor sessions and additional support;
- number of logins to the online learning platform;
- evidence of changes to teaching including use of formative assessment processes and/or developmental progressions;
- use of maths practice scale, the Behaviour for Learning scale, website materials, and home learning materials; and
- evidence of nominated Reception and Year 1 teaching working together.

Child recruitment rate and child assessment completion rates were also recorded by the evaluation team.

An overview of the IPE design and methods is provided in Table 9.

Table 9: IPE research questions, logic model relevance, and methods of data collection and analysis overview

Research questions	Implementation / logic model relevance	Data collection methods	Participants / data sources	Data analysis
4	Usual practice; monitoring practice	Usual practice surveys: Online questionnaires (paper where needed)	Nominated Reception/Year 1 teachers Staff member with oversight of maths teaching and CPD plans All intervention and control schools	Descriptive statistics; content analysis
	Understand delivery and the intervention	Observations	PD training workshops; attending Reception/Year 1 teachers and senior management; the delivery team Sessions 1, 2, 7, and 8	Inductive coding; thematic analysis; cross-case analysis
1, 2, 3, 4, 6, 7	Stakeholders' viewpoints on the intervention, fidelity, implementation, programme adaptations, differentiation, and barriers/facilitators to implementation, understand how on-treatment status may impact on outcomes; understanding any transfer of learning to the home	Semi-structured interviews	Participating and non-participating Reception/Year 1 teachers and teaching assistants; senior management; school's allocated TEEMUP PD mentor	Inductive coding; thematic analysis

	learning environment, collection of cost data			
		Observations	Observation of in-school mentoring session involving mentor and relevant teachers*	Cross-case analysis
1, 2, 3, 4	Gather stakeholders' viewpoints on the intervention programme adaptations, differentiation, and barriers/facilitators to implementation, understanding any transfer of learning to the home learning environment, collection of cost data, usual practice (where applicable)	Surveys: Online questionnaires, (paper where needed)	All nominated Reception/Year 1 in all intervention schools	Descriptive statistics; content analysis
1, 2, 5, 7	Fidelity, Participant responsiveness, Dosage	Monitoring data: Training attendance registers, mentor visit logs, mentor proformas, website analytics, trial recruitment and attrition rates	All intervention; control schools where relevant	Descriptive statistics; cross-case analysis
1, 2, 6, 7	Delivery, implementation, perceived impact, including actual barriers to and facilitators of delivery/implementation, future plans, adaptation of the programme, readiness for effectiveness trial, validation of logic model	Semi-structured interviews	Delivery team	Inductive coding; thematic analysis

**Data on school's experience of the TEEMUP mentor sessions collected via surveys instead of observations.*

Analysis

Interviews were transcribed verbatim. Data organisation was facilitated through the use of NVIVO. For case studies, each school was treated as a case site and an overview is provided in relation to their usual practice, delivery, fidelity, and implementation of the intervention.

To achieve a systematic approach to data analysis, all interviews were analysed thematically following the stages outlined by Braun and Clarke (2006): detailed familiarisation; generating initial codes; searching for themes; reviewing themes; defining and naming themes; and data reporting. Initially the transcripts were inductively analysed to develop a broad coding framework. The development of the coding framework was iterative and refined throughout analysis and used as a framework for remaining transcripts to be coded.

Surveys were analysed using content analysis, with descriptive statistics presented.

Results from all IPE data collection methods were synthesised from the themes and presented as answers to the IPE research questions. Cross-case analysis describes themes, similarities, and differences found within case study interviews and results presented where appropriate alongside other relevant IPE data. Interview and survey quotes were chosen and included within the report as they best represent the essence of the identified themes; they have meaning, capture key aspects, and provide strong evidence of the interpretation of the data. Findings from the IPE were used to explore the programme theory and relationships between different components of the logic model with an aim to test its causal assumptions and inform intervention scale-up. The IPE findings can help to explain the results of the impact evaluation TEEMUP PD intervention on children's maths outcomes.

Costs evaluation design

The cost analyses followed the 'ingredients method' (Levin *et al.,* 2018) to account for all costs of intervention implementation.

Sources of data

Participating schools were requested to complete surveys to capture usual practice at baseline (Autumn Term 2021). Follow-up surveys collected data from nominated Reception and Year 1 teachers in intervention schools on implementation and costs associated with delivering the TEEMUP PD at the time points: mid-implementation (Summer Term 2022), and at the end of the trial (Summer Term 2023).

The total number of hours that teachers spent completing workshops in each calendar year was obtained from registers of attendance recorded by the delivery team.

June and July 2022 (related to the academic year 2021–2022)

This survey captured the number of hours in total the teacher spent meeting with the TEEMUP mentor (online, over the phone, via email, or in person) and completing other TEEMUP activities. The survey also captured whether, and for how many hours, the school had to arrange for paid cover for the teacher to complete these activities, and the cost of travel to the venue(s) to attend in-person TEEMUP workshops. If staff indicated they had spent time completing TEEMUP PD workshops and/or related activities outside of normal working schools' hours/in their own time, the survey asked for a breakdown of time spent on each activity. The survey also requested information on what existing resources, if any, the teacher had used as a result of implementing TEEMUP PD so far (e.g. computer, internet connection, learning resources) and any additional resources, with the quantity and cost, they had bought.

June and July 2023 (related to the academic year 2022–2023)

This survey captured the same data as the Summer Term 2022 survey but asked about the academic year 2022–2023.

Cost data was also collected directly from the delivery team in order to estimate the cost of delivering the intervention.

Analysis

The total time devoted by school personnel for training and intervention delivery/engagement are summarised descriptively. The average cost to the school of taking part in the intervention was calculated by summing the constituent parts (cost of teacher cover, travel to training venues, and purchase of additional resources) and dividing by the number of schools who responded to the surveys.

Per-pupil costs were determined by summing the total costs per school and dividing by the average number of pupils per school eligible for inclusion in this evaluation (n=50, 25 per year group). This figure is higher than the average number of pupils actually included in the evaluation as this was limited to approximately 15 per year group to reduce the cost of implementing outcome assessments.

The cost to the Oxford team of the development, set-up, and delivery of the intervention are summarised and presented as an average cost per school and an average cost per pupil school year across three years.

Assumptions

- Data for the costs are based on responses to the surveys. Not all schools provided data and different schools provided data at the different time points. Additionally, not all teachers within the schools that were represented may have provided data. However, for the purposes of these analyses, we have to make the assumption that the teachers and schools responding to the surveys represent the whole sample of

intervention schools and so the costs and time commitments reported by them are generalisable to all randomised schools.

- For the purpose of calculating the costs over a three-year period, we assumed that the costs incurred by schools of staff cover, travel to attend face-to-face workshops and the purchase of additional resources captured in the final survey (about the academic year 2022–2023) would only be required in Year 1. Some materials may need to be replaced annually or additional, similarly priced materials may need to be purchased, but these costs are minimal and are considered as optional extras.

- For the purposes of calculating the average cost of paid cover for teachers to complete workshops, meet their mentor, and undertake additional activities relating to the TEEMUP intervention, we assumed that the cost of staff cover is £18.21 per hour, which is a representative hourly rate of a primary school teacher in the UK.[3]

[3] The pay for classroom teachers in England, excluding London, ranged from £25,714 to £36,961 across six spine points in 2021–2022 according to the National Association of Schoolmasters Union of Women Teachers (NASUWT), see **here**. Taking £30,775 as a midpoint between the middle two spine points and averaging over 52 weeks a year and 32.5 hours a week equals £18.21.

Timeline

Table 10: TEEMUP timeline

Dates	Activity	Staff responsible / leading
November 2019 – January 2020	• Set-up	All
January 2020 – November 2020	• Ethical application	Evaluation team
March 2020 – January 2021	• Disruption due to COVID-19 (the trial start date in schools was delayed by one year; for more details, see the 'Impact evaluation design' section of this report)	Not applicable
January 2021 – October 2021	• School recruitment	Delivery team
September 2021 – December 2021	• Disseminate trial information to parents/carers of Cohort 1 • Schools provide pupil participation lists	Evaluation team
October 2021 – December 2021	• Cohort 1 baseline data collection (BAS3 ENC, CSBQ) • Confidence surveys (baseline) • IPE usual practice survey (baseline)	Evaluation team
November 2021 – December 2021	• Batch randomisation	Evaluation team
January 2022 – May 2023	• TEEMUP PD programme delivery and collection of monitoring data	Delivery team
May 2022 – July 2023	• IPE interviews (longitudinal)	
June 2022 – July 2022	• Cohort 1 CSBQ data collection (midpoint) • Staff confidence surveys (midpoint)	Evaluation team
June 2022 – July 2022	• IPE surveys (intervention group only)	Evaluation team
December 2022	• IPE delivery team interview	Evaluation team and delivery team
October 2022 – January 2023	• Disseminate trial information to parents/carers of Cohort 2 • Schools provide participation lists	Evaluation team
November 2022 – January 2023	• Cohort 2 CSBQ data collection (baseline)	Evaluation team
October 2022 – December 2022	• IPE usual practice survey (midpoint)	Evaluation team
September 2022 – July 2023	• IPE interviews	Evaluation team
June 2023 – July 2023	• Post-test assessments with Cohort 1 and Cohort 2 pupils (BAS3 ENC, CSBQ)	Evaluation team
June 2023 – July 2023	• IPE usual practice surveys (Endpoint) • Staff confidence surveys (Endpoint) • IPE post-surveys (intervention delivery)	Evaluation team
August 2023	• Submission of NPD request for EYFSP data for Cohort 1 and Cohort 2	Evaluation team
September 2023	• IPE delivery team interview	Evaluation team and delivery team
August 2023 – February 2024	• Data analysis • Report writing	Evaluation team
February 2024	• Submit draft report to the EEF	Evaluation team
June 2024	• Final report and data uploaded to the EEF data archive • Updating of ISRCTN trial registry with results • Submission of final statement of spend to the EEF	Evaluation team

Impact evaluation results

Participant flow including losses and exclusions

School recruitment and attrition

Between January 2021 and September 2021, 2,356 schools were approached to take part in the trial, of which 206 (8.7%) returned an Expression of Interest form. MoUs were sent to 177 eligible schools and were signed and returned by 101 schools. Reasons MoUs were not sent to 29 schools that returned an Expression of Interest form include: another school from the same MAT already returned an MoU (*n*=12); the school cannot commit to keeping staff in relevant year groups (*n*=7); participating in other research (*n*=3); duplicate application (*n*=3); Reception class too small (*n*=3); and out of area (*n*=1). Six schools that returned an MoU withdrew from the trial before they were randomised; for four of which this was before pupil details were requested from the school (school stopped engaging with the evaluation team [*n*=3] or had other commitments [*n*=1]). Pupil details were requested from 97 schools and were returned from 95 schools following which baseline testing was arranged to take place. Two schools could not accommodate a baseline assessment visit citing disruption due to the COVID-19 pandemic as a reason. In total, 93 schools were randomised into the trial following baseline data collection; 47 to the intervention group and 46 to control. Following randomisation, one intervention school fully withdrew from the evaluation prior to CSBQ post-testing at the end of the first academic year for Cohort 1, as they felt it was not a good time for their school to be involved in the trial.

At the point of recruiting Cohort 2, a control school fully withdrew from the evaluation due to staffing pressures within their school. Furthermore, 18 schools declined to recruit a Cohort 2 and therefore, did not provide pupil details for their eligible pupils who started in Reception in September 2022. These 18 schools agreed to continue with the evaluation for Cohort 1, though nine of these schools only agreed to accommodate post-testing for the primary outcome (i.e. BAS3 ENC) and declined to be involved in other evaluation activities, such as completing the CSBQ for the final follow-up. Of the 18 schools that withdrew from Cohort 2, 14 schools explained that this was due to staff changes, staffing issues, or staff capacity. The four other schools did not provide a reason and did not provide pupil details for Cohort 2 after several contact attempts from the evaluation team.

Cohort 1: Child recruitment and attrition

For Cohort 1, at the start of the academic year 2021–2022, interested schools were asked to confirm the number of pupils in Reception at the school who were aged four to five years and being taught by the nominated Reception teacher. These data were requested from the 101 schools that returned an MoU and were returned by 96 schools. These 96 schools identified a total of 2,369 eligible pupils (mean per school 24.7, SD 5.9, median 26, range 10 to 40).

Among the schools for whom pupil details were requested, information sheets and withdrawal forms were sent to the parents/carers of these identified pupils. Pupils for whom a withdrawal form was not received were considered eligible for baseline assessment. A total of 95 schools returned pupil details, of which 70 also indicated how many pupils they had received a withdrawal form for or who were ineligible based on having significant SEND or EAL where an extreme language barrier existed or the pupil was new to English, which would have prevented them from accessing the assessment and/or cause distress through completing the assessment. The 70 schools indicated that a total of 68 withdrawal forms had been returned (mean per school 1.0, SD 1.8, median 0, range 0 to 10) and 94 pupils (from 65 schools) were deemed ineligible (mean per school 1.4, SD 1.7, median 1, range 0 to 6).

In total, among the 93 randomised schools, 2,192 pupils were considered eligible for baseline assessment. This list underwent the sampling exercise described in the 'Methods' section to select and order pupils for assessment with the BAS3 ENC. A total of 1,905 pupils were selected (mean per school 20.5, SD 2.8, median 22, range 11 to 22) to try and obtain BAS3 ENC baseline data for. A further 287 were considered as 'Optional' for testing for the CSBQ at baseline.

The primary outcome was the BAS3 ENC score for Cohort 1 and we defined the randomised sample as being all pupils with a valid baseline BAS3 ENC score (*n*=1,566; intervention, *n*=780; control, *n*=786; 93 schools, mean per school 16.8, SD 2.1,

median 18, range 9 to 22). The item at which pupils started the BAS3 ENC assessment was dependent on their age at the time; 436 started at item 1 and 1,130 at item 4.

Post-testing with the BAS3 ENC was due to take place for 1,488 pupils with valid baseline data across 91 schools (intervention, *n*=745/780, 95.5%; control, *n*=743/786, 94.5%). Post-testing was not due to take place with 78 pupils as the evaluation team had been informed that they had left the school (*n*=49) or their school had decided to withdraw from the trial (*n*=29 pupils from two schools [intervention, *n*=1; control, *n*=1]) before this was due.

A post-test BAS3 ENC score was avalable for 1,363 pupils from 91 schools; however, for two of these, post-testing was not expected as the child had not completed a valid BAS3 ENC baseline assessment. Therefore, a valid follow-up BAS3 ENC score was avalable for 1,361 pupils from 91 schools (intervention, *n*=679/780, 87.1%; control *n*=682/786, 86.8%; mean per school 15.0, SD 2.4, median 15, range 4 to 20). Reasons for data being unavailable for the remaining 127 include: child had left school (*n*=81), child was absent on day of testing (*n*=35), and other reasons (*n*=11).

The item at which pupils started the BAS3 ENC post-test was dependent on their age at the time; 914 started at item 4 and 447 at item 18.

Baseline CSBQ was obtained for 1,954 pupils (intervention, *n*=975; control, *n*=979) across all 93 randomised schools (mean per school 21.0, SD 4.3, median 21, range 11 to 31).

CSBQ post-test scores were provided at the first follow-up for 1,170 pupils, of which 1,156 had a valid baseline score, from 63 schools (31 in the intervention group, and 32 in control; mean per school 18.3, SD 5.2, median 18, range 1 to 29).

CSBQ post-test scores were provided at the second follow-up for 1,372 pupils, of which 1,368 had a valid baseline score, from 72 schools (38 in the intervention group, and 34 in control; mean per school 19.0, SD 4.0, median 19, range 8 to 30).

Cohort 2: Child recruitment and attrition

At the start of the Autumn Term 2022, the evaluation team requested pupil details for Cohort 2 from schools that were still engaging with the project. These were received from 73 schools since two had fully withdrawn before then and a further 18 had withdrawn from Cohort 2. Details were provided for 1,768 pupils (mean per school 24.2, SD 5.9, median 26, range 10 to 38), of which 1,084 were sampled for assessment (mean per school 14.8, SD 1.1, median 15, range 9 to 17).

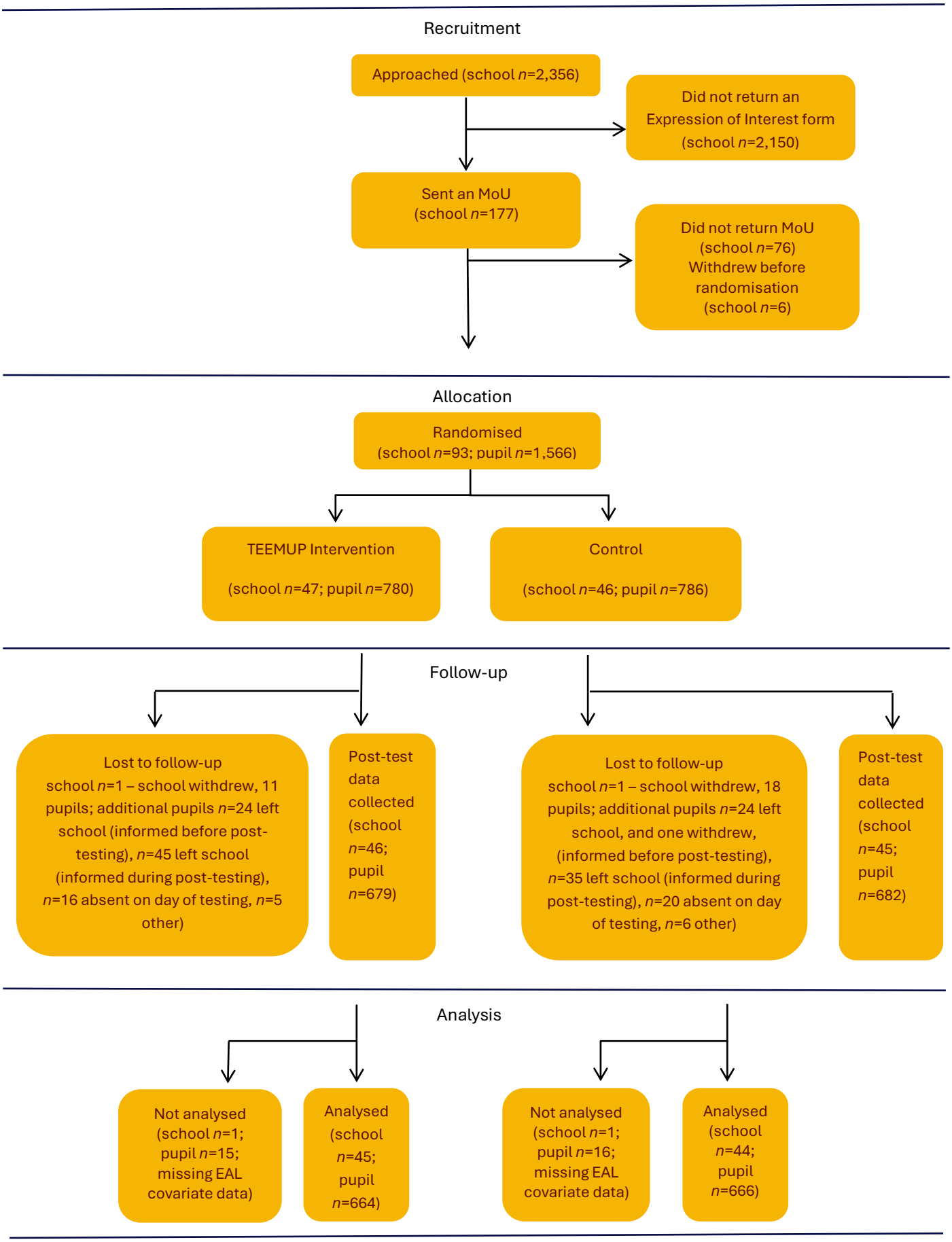
Baseline testing with the BAS3 ENC at the start of the academic year 2022–2023 was not conducted for Cohort 2; the BAS3 ENC was only completed once for outcome assessment during June 2023 to July 2023. Baseline CSBQ was, however, obtained for 969 pupils (intervention, *n*=484; control, *n*=485) across 67 schools (34 in the intervention group and 33 control; mean per school 14.5, SD 2.0, median 15, range 1 to 15). Six schools did not provide CSBQ baseline data; four of these only returned pupil details towards the very end of the data collection period for baseline CSBQ so a decision was made not to ask these schools to complete the baseline CSBQ for Cohort 2 at that stage and instead just to ask them to complete the post-test, and two did not return CSBQ baseline data despite being asked several times by the evaluation team.

CSBQ post-test scores were provided for 928 pupils (intervention, *n*=475; control, *n*=453) from 64 schools (33 in the intervention group, and 31 in control), of which 866 (intervention, *n*=430; control, *n*=436), from 61 schools (31 in the intervention group, and 30 in control), had a valid baseline score (mean per school 14.2, SD 2.1, median 15, range 1 to 15). Six schools that provided baseline CSBQ data did not provide post-test CSBQ data, and three schools that did not provide CSBQ data at baseline provided it at post-test.

A valid follow-up BAS3 ENC score was avalable for 1,001 pupils (intervention, *n*=525; control *n*=476; mean per school 13.7, SD 1.5, median 14, range 9 to 15) from 73 schools (38 in the intervention group, and 35 in control). The item at which pupils started the BAS3 ENC post-test was dependent on their age at the time; all started at item 4.

Figure 2 depicts the flow of schools and pupils through the trial.

Figure 2: Participant flow diagram (two arms)



As stated in the protocol, we aimed to have 80% power to detect an effect size of 0.21 of an SD with 1,500 pupils from 100 schools, assuming a baseline post-test correlation of 0.60, an ICC at the school level of 0.15, and 15% loss to follow-up (Table 11). At randomisation, we had valid baseline BAS3 ENC data for 1,566 pupils and, under otherwise identical assumptions, the MDES was 0.22. The actual observed ICC at the school level obtained from the primary analysis model was 0.04 (95% CI: 0.02 to 0.07). The overall correlation between the pre- and post-test scores of the participants included in the primary analysis was 0.47. Based on the number of pupils included in the primary analysis model (*n*=1,330), and the observed ICC and baseline post-test correlation, the estimated MDES at analysis for the primary outcome was 0.17 (Table 11).

Within the subgroup of randomised pupils eligible for FSM (*n*=303, intervention, *n*=160/780, 20.5%; control, *n*=143/786, 18.2%), 238 pupils were analysable (i.e. had valid BAS3 ENC follow-up data and data for the covariates specified for the primary analysis). The observed ICC at the school level obtained from the restricted primary analysis model was 0.00 (95% CI: 0.00 to 0.00) and the correlation between the pre- and post-test scores was 0.40. Hence, the estimated MDES at analysis for the primary outcome among the subgroup eligible for FSM was 0.33 (Table 11).

Table 11: MDES at different stages

		Protocol[a]		Randomisation[a]		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.21	0.31	0.22	0.32	0.17	0.33
Pre-/post-test correlations	Level 1 (pupil)	0.60	0.60	0.60	0.60	0.47	0.40
ICCs	Level 2 (school)	0.15	0.15	0.15	0.15	0.04	0.00
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two	Two	Two
Average cluster size		15	3	17	3.7	14.9	3.2
Number of schools	Intervention	50	50	47	41	45	37
	Control	50	50	46	40	44	37
	Total:	100	100	93	81	89	74
Number of pupils	Intervention	750	150	780	160	664	126
	Control	750	150	786	143	666	112
	Total:	1,500	300	1,566	303	1,330	238

[a]*MDES assumes 15% loss to follow-up.*

Attrition

A total of 1,361 of the 1,566 randomised pupils (intervention, *n*=679/780, 87.1%; control *n*=682/786, 86.8%) from 91 schools had valid baseline and post-test BAS3 ENC data. However, 31 did not have data for the EAL covariate, so 1,330 from 89 schools were included in the primary analysis. This equates to an overall attrition rate of 15.1% (Table 12).

Table 12: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	780	786	1,566
	Analysed	664	666	1,330
Pupil attrition (from randomisation to analysis)	Number	116	120	236
	Percentage	14.9	15.3	15.1

Pupil and school characteristics

Characteristics for the 93 randomised schools and 1,566 participating pupils in Cohort 1 are presented in Table 13 and appear broadly similar between the two groups, except that there is a higher proportion of schools in an urban locale in the control group than intervention (67.4% vs 53.2%). No formal hypothesis testing was performed on baseline data (Senn, 1994), so comparisons are made visually only.

At enrolment to the trial, schools allocated to the intervention group had, on average, 20.0% of pupils eligible for FSM and 18.2% of pupils with EAL (i.e. first language known or believed to be other than English). Similarly, the control group had, on average, 18.0% of pupils eligible for FSM and 17.7% EAL pupils (Table 13).

Approximately half of the participating pupils were male (intervention, *n*=384/780, 49.2%; control, *n*=382/786, 48.6%) and, overall, 19.4% were eligible for FSM (intervention, *n*=160/780, 20.5%; control, *n*=143/786, 18.2%). A slightly higher proportion of pupils in the control group than the intervention group had EAL (18.4% vs 13.8%). There was a small amount of missing data in this variable (intervention, *n*=16/780, 2.1%; control, *n*=18/786, 2.3%).

The best available source we could identify for providing some national-level data to assess the representativeness of our recruited sample of schools and pupils is a government report published annually on schools, pupils, and their characteristics, the latest publication being June 2023.[4] Data indicate that the percentage of pupils in state-funded primary schools ever eligible for FSM in the last six years in 2021–2022 was 25.5%, slightly higher than the average percentage of ever FSM pupils in our 93 recruited schools at the time of enrolment into the trial, 19.0%. Similarly, 18.8% of pupils in Reception of state-funded primary schools, nationally, were currently eligible for FSM in the year 2021–2022, which is the year we recruited the Reception cohort. Our randomised sample for Cohort 1 had a similar proportion of pupils eligible for FSM (19.4% overall). Nationally, in state-funded primary schools in England 21.2% of pupils were known or believed to have EAL in 2021–2022 and 22.0% in 2022–2023. Within our sample, schools reported an average of 17.9% of their pupils having EAL.

Another source of national data is the 'Get Information about Schools' government service. According to data accessed in October 2022, restricting to academies and local authority maintained schools providing primary phase education in England, just under half are community schools (47.6%), 17.0% are academy converter, 15.2% are voluntary aided schools, 10.2% are voluntary controlled school, 6.3% are academy sponsor-led, and 3.7% are foundation schools. Three-quarters

[4] https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics

are in urban locales (73.1%), 25.2% are rural, and 1.7% are missing these data. An Ofsted (Office for Standards in Education, Children's Services and Skills) rating was unavailable for 38.7% of schools, but for those that did have these data, 15.4% were rated 'Outstanding' in their latest inspection, 75.6% 'Good', and 9.0% 'Requires improvement'.

In our sample, relative to the national picture, we have a smaller proportion of community schools (35.5%) and a slightly higher proportion of academy converter schools (22.6%); we have a smaller proportion of schools in urban areas and a higher proportion in rural areas (60.2% and 40.2%, respectively); and, of the schools for which the latest Ofsted inspection rating was available, we have a lower proportion of schools rated 'Outstanding' (9.1%) and a higher proportion of 'Good' (83.0%).

The average baseline BAS3 ENC score was 44.7 (SD 7.4) out of a possible 80 in the intervention group and 44.9 (SD 7.8) in the control group (Hedges' g effect size between the groups -0.03; 95% CI: -0.13 to 0.07; Table 13).

Characteristics for the 89 schools and 1,330 pupils included in the primary analysis are presented in Table 14, and are very similar to the randomised population.

Table 13: Baseline characteristics of randomised schools and pupils, by group, as randomised

School level (categorical)	Intervention group (*n*=47)	Control group (*n*=46)	
	Count (%)	Count (%)	
Location			
Milton Keynes	10 (21.3)	10 (21.7)	
Newmarket/Bury St Edmunds	9 (19.1)	10 (21.7)	
Oxford	9 (19.1)	8 (17.4)	
Norwich	7 (14.9)	6 (13.0)	
Peterborough	6 (12.8)	6 (13.0)	
Barnet	6 (12.8)	6 (13.0)	
%FSM[a]			
≤16%	22 (46.8)	24 (52.2)	
>16%	25 (53.2)	22 (47.8)	
%EAL[a]			
≤8%	23 (48.9)	24 (52.2)	
>8%	24 (51.1)	22 (47.8)	
Rural urban classification			
Urban	25 (53.2)	31 (67.4)	
Rural	22 (46.8)	15 (32.6)	
Type of school			
Community school	15 (31.9)	18 (39.1)	
Academy converter	9 (19.1)	12 (26.1)	
Voluntary controlled school	7 (14.9)	7 (15.2)	
Voluntary aided school	8 (17.0)	4 (8.7)	

Academy sponsor-led or Foundation school[b]	8 (17.0)	5 (10.9)	
Ofsted rating[b]			
Outstanding or Good	41 (87.2)	40 (87.0)	
Requires improvement or Missing	6 (12.8)	6 (13.0)	
Participation in NCETM	25 (53.2)	28 (60.9)	
School level (continuous)	Mean (SD)	Mean (SD)	
%FSM[a]	20.0 (10.2)	18.0 (9.4)	
%EAL[a]	18.2 (22.4)	17.7 (20.2)	
Cohort 1	**N=780**	**N=786**	
Pupil level (categorical)	Count (%)	Count (%)	
Gender			
Male	384 (49.2)	382 (48.6)	
Female	396 (50.8)	404 (51.4)	
Eligible for FSM	160 (20.5)	143 (18.2)	
EAL	108 (13.8)	145 (18.4)	
Pupil level (continuous)	Mean (SD)	Mean (SD)	Effect size
Age (years)	4.3 (0.4)	4.3 (0.4)	—
BAS3 ENC	44.7 (7.4)	44.9 (7.8)	-0.03 (-0.13, 0.07)
CSBQ Sociability	3.8 (0.8)	3.6 (0.9)	0.18 (0.08, 0.28)
CSBQ Externalising problems	1.5 (0.7)	1.6 (0.9)	-0.11 (-0.21, -0.01)
CSBQ Internalising problems	1.5 (0.6)	1.6 (0.7)	-0.13 (-0.22, -0.03)
CSBQ Prosocial behaviour	3.8 (0.8)	3.7 (0.9)	0.19 (0.09, 0.29)
CSBQ Behavioural self-regulation	4.0 (0.9)	3.8 (0.9)	0.16 (0.06, 0.26)
CSBQ Cognitive self-regulation	3.4 (0.9)	3.3 (0.9)	0.12 (0.02, 0.22)
CSBQ Emotional self-regulation	4.1 (0.7)	3.9 (0.8)	0.20 (0.10, 0.30)
CSBQ Composite self-regulation	3.8 (0.7)	3.7 (0.8)	0.19 (0.09, 0.28)

[a]Based on data provided by schools at the start of the evaluation, dichotomised for the randomisation at the median level for the recruited schools. FSM based on % of pupils eligible for FSM at any time during the past six years.
[b]Rows combined to prevent statistical disclosure relating to cell counts for schools <4.

Table 14: Baseline characteristics of randomised schools and pupils, by group, as included in the primary analysis

School level (categorical)	Intervention group (*n*=45)	Control group (*n*=44)	
	Count (%)	Count (%)	
Location			
Milton Keynes	10 (22.2)	10 (22.7)	
Newmarket/Bury St Edmunds	9 (20.0)	10 (22.7)	

Oxford	9 (20.0)	8 (18.2)	
Norwich	7 (15.6)	5 (11.4)	
Peterborough	6 (13.3)	5 (11.4)	
Barnet	4 (8.9)	6 (13.6)	
%FSM[a]			
≤16%	22 (48.9)	23 (52.3)	
>16%	23 (51.1)	21 (47.7)	
%EAL[a]			
≤8%	22 (48.9)	23 (52.3)	
>8%	23 (51.1)	21 (47.7)	
Rural urban classification			
Urban	24 (53.3)	29 (65.9)	
Rural	21 (46.7)	15 (34.1)	
Type of school			
Community school	15 (33.3)	17 (38.6)	
Academy converter	9 (20.0)	12 (27.3)	
Voluntary controlled school	7 (15.6)	6 (13.6)	
Voluntary aided school	6 (13.3)	4 (9.1)	
Academy sponsor-led or Foundation school[b]	8 (17.8)	5 (11.4)	
Ofsted rating[b]			
Outstanding or Good	39 (86.7)	38 (86.4)	
Requires improvement or missing	6 (13.3)	6 (13.6)	
Participation in NCETM	24 (53.3)	26 (59.1)	
School level (continuous)	Mean (SD)	Mean (SD)	
%FSM[a]	19.5 (10.1)	18.1 (9.5)	
%EAL[a]	17.4 (21.6)	18.2 (20.5)	
Cohort 1	**N=664**	**N=666**	
Pupil level (categorical)	Count (%)	Count (%)	
Gender			
Male	330 (49.7)	328 (49.2)	
Female	334 (50.3)	338 (50.8)	
Eligible for FSM	126 (19.0)	112 (16.8)	
EAL	93 (14.0)	122 (18.3)	
Pupil level (continuous)	Mean (SD)	Mean (SD)	Effect size
Age (years)	4.3 (0.4)	4.3 (0.4)	—

BAS3 ENC	44.9 (7.7)	45.1 (7.8)	-0.03 (-0.14, 0.08)
CSBQ Sociability	3.8 (0.8)	3.7 (0.8)	0.19 (0.08, 0.29)
CSBQ Externalising problems	1.5 (0.8)	1.6 (0.9)	-0.10 (-0.21, 0.01)
CSBQ Internalising problems	1.5 (0.6)	1.6 (0.7)	-0.14 (-0.25, -0.03)
CSBQ Prosocial behaviour	3.9 (0.8)	3.7 (0.9)	0.21 (0.1, 0.31)
CSBQ Behavioural self-regulation	4.0 (0.9)	3.8 (0.9)	0.18 (0.07, 0.28)
CSBQ Cognitive self-regulation	3.5 (0.9)	3.3 (0.9)	0.15 (0.04, 0.25)
CSBQ Emotional self-regulation	4.1 (0.7)	3.9 (0.8)	0.20 (0.10, 0.31)
CSBQ Composite self-regulation	3.8 (0.7)	3.7 (0.8)	0.20 (0.10, 0.31)

[a]*Based on data provided by schools at the start of the evaluation, dichotomised for the randomisation at the median level for the recruited schools.*
[b]*Rows combined to prevent statistical disclosure relating to cell counts for schools <4.*

Outcomes and analysis

Primary analysis

Cohort 1

A valid baseline BAS3 ENC score was obtained for 1,566 pupils (intervention, *n*=780; control, *n*=786), from all 93 randomised schools. The ICC associated with school for the baseline score is 0.09 (95% CI: 0.05 to 0.14). In total, a valid post-test BAS3 ENC score was available for 1,361 randomised pupils (86.9%; intervention, *n*=679/780, 87.1%; control, *n*=682/786, 86.8%), from 91 schools (46 intervention, 45 control). A mean of 48.4 (95% CI: 47.5 to 49.2) was observed in the intervention arm and 48.1 (95% CI: 47.3 to 49.0) in the control arm. The unadjusted mean difference is 0.21 (95% CI: -0.96 to 1.38) (Appendix D Table 1). Histograms of the baseline and post-test scores show that pre-test data are relatively normally distributed while the post-test data have a non-symmetric distribution (Figure 3). The correlation between the baseline and post-test scores was 0.47 (95% CI: 0.43 to 0.51). While baseline and post-test scores were available for 1,361 pupils, 31 did not have EAL covariate data and so could not be included in the analysis model. As a check of the analysis model assumptions, the normality of the standardised residuals was checked using a QQ plot and was shown to be normally distributed (Figure 4). The adjusted mean difference in post-test score between the intervention and control groups was 0.73 (95% CI: -0.55 to 2.02), p = 0.26, Appendix D Table 1). The estimated Hedges' g effect size was 0.07 (95% CI: -0.05 to 0.18), which relates to one additional month's progress in the intervention group (Table 15). The total variance used to calculate the effect size was 121.26; the sum of 111.21 (random variation between pupils, within-cluster variance) and 10.05 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.04 (95% CI: 0.02 to 0.07). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.08 (0.06 to 0.12). The Stata analysis code to produce the results for the primary outcome are included in the Technical Notes for this report.

Figure 3: Histograms of baseline and post-test BAS3 ENC scores (scores have been aggregated to present the mean score for every ten pupils to prevent statistical disclosure)

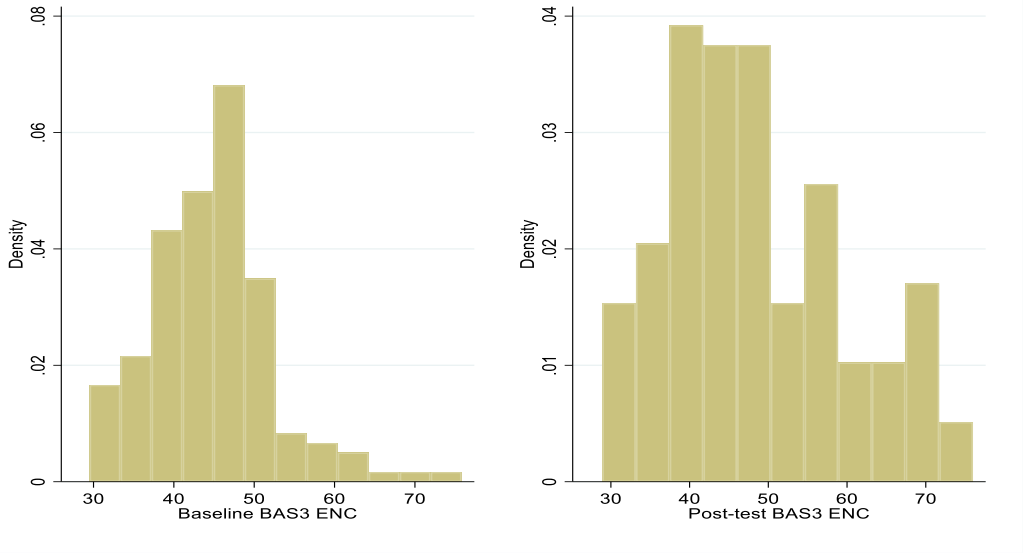


Figure 4: QQ plot of the standardised residuals from the primary analysis model to assess the normality assumption (residuals have been aggregated to present the mean residual for every ten pupils to prevent statistical disclosure)[5]

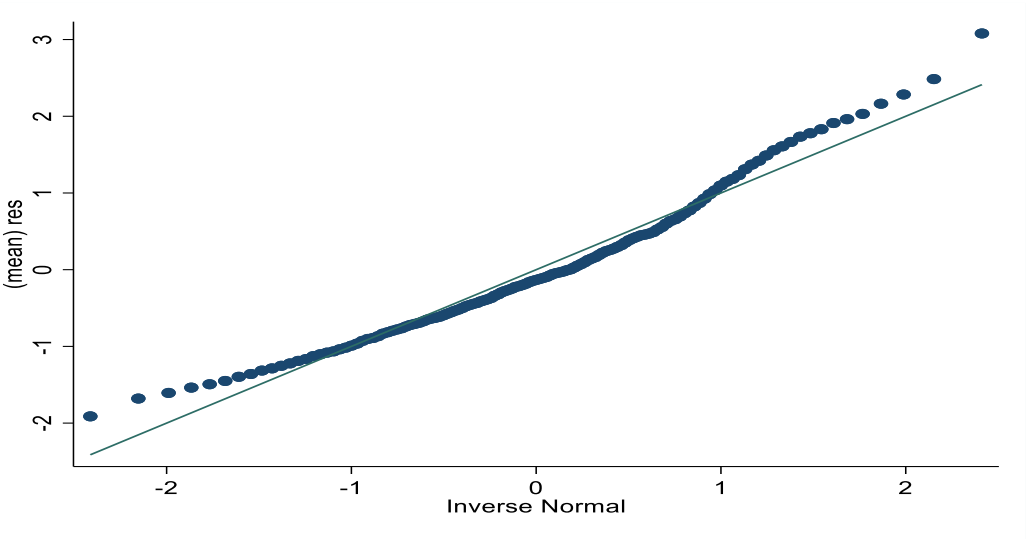


Table 15: Primary analysis (primary outcome)

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	*n* (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
BAS3 ENC	679 (101)	48.4 (47.5, 49.2)	682 (104)	48.1 (47.3, 49.0)	1,330 (664; 666)	0.07 (-0.05, 0.18)	0.26

[5] QQ plot demonstrates how the distribution quantiles of our model residuals (y axis) line up with the normal distribution (x axis)—the dots should not deviate too far from the reference line.

Subgroup analyses

Summary statistics for the post-test BAS3 ENC score are presented by gender, FSM eligibility, and whether schools were taking part in the NCETM Maths Hubs Programme (Table 16). These summaries indicate that, in general, pupils eligible for FSM performed slightly worse on the post-test than those not eligible for FSM. Scores were reasonably similar for male and female pupils, and for pupils stratified by whether their school participated in the NCETM Maths Hubs Programme.

In a series of adjusted regression analyses that included interaction effects, there was no evidence of an interaction between trial allocation and: i) gender (interaction term coefficient 1.30, 95% CI: -0.80 to 3.40, p = 0.22); ii) FSM eligibility (0.94, 95% CI: -1.72 to 3.60, p = 0.49); or iii) school participation in NCETM Maths Hubs Programme (2.55, 95% CI: -0.12 to 5.21, p = 0.06).

Within the randomised sample, there was 303 pupils from 93 schools who were eligible for FSM (median per school 2, range 0 to 11 in both groups). Within the restricted sample of pupils eligible for FSM, the correlation between the baseline and post-test scores was 0.40 (95% CI: 0.29 to 0.50). The adjusted mean difference in post-test score between the intervention and control groups within this sample was 1.30 (95% CI: -1.03 to 3.63, p = 0.27) and the estimated Hedges' g effect size was 0.13 (95% CI: -0.10 to 0.37), which relates to two additional months' progress in the intervention group. The total variance used to calculate the effect size (from a model without covariates) was 97.92—the sum of 97.92 (random variation between pupils, within-cluster variance) and 0.00 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.00 (95% CI: 0.00 to 0.00). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.00 (0.00 to 0.00).

As a sensitivity check, the effect size for the FSM-eligible subgroup from the interaction effects model was extracted and compared with the effect size derived from the restricted sample. The adjusted mean difference in post-test score between the intervention and control groups within this sample was 1.51 (95% CI: -0.88 to 3.89, p = 0.22) and the estimated Hedges' g effect size was 0.14 (95% CI: -0.08 to 0.35), which relates to two additional months' progress in the intervention group.

Table 16: Subgroup summary scores for the post-test BAS3 ENC score

Outcome	Raw means			
	Intervention		Control	
	n	Mean (SD)	*N*	Mean (SD)
Gender				
Male	336	49.8 (11.4)	336	49.3 (11.8)
Female	343	46.9 (10.6)	346	47.0 (9.9))
FSM eligibility				
Yes	133	45.2 (10.7)	113	44.3 (9.0)
No	546	49.1 (11.1)	569	48.9 (11.1)
School's participation in NCETM Maths Hubs Programme				
Yes	352	49.5 (11.1)	421	48.0 (11.1)
No	327	47.1 (11.0)	261	48.3 (10.7)

Analysis in the presence of non-compliance

Summaries of compliance

Nine intervention schools withdrew from implementing the intervention and were automatically classed as 'non-compliant' for both Cohorts 1 and 2. Three withdrew following randomisation, prior to delivery of any PD training, one withdrew part-

way through PD training workshop delivery, and five withdrew in the autumn of the second year of the trial. Reasons cited were largely related to staff changes and absences, and time and financial constraints, including the cost of covering for staff to attend the PD sessions.

More detailed compliance data were collected for the remaining 38 intervention schools and is summarised in Table 17. Overall, including the nine schools that withdrew from the intervention, 22 of the 47 intervention schools (46.8%) were classed as having 'good' compliance for Cohort 1, eight (17.0%) as 'minimal', and 17 (36.2%) as non-compliant. The delivery team became aware of one instance during the trial of a control school joining a MAT that contained a trial intervention school and, anecdotally, that some teachers in the control school may have received some cascade training from teachers in the intervention school. However, according to our pre-specified definitions of intervention compliance, this control school would not have received any element of the intervention sufficiently to even be classed as a 'minimal' complier, for example, none of their teachers attended any training sessions. Therefore, all 46 control schools were classed as non-compliant.

In total, 36 of the 38 intervention schools had a trained Reception teacher (based on attending at least seven out of nine PD sessions) who remained teaching Reception for the majority of the year (2021–2022). (Two of the schools were reported as having a change in Reception teacher part-way through Reception so the data for the teacher who taught for most of the year was used.) A total of 33 of the 38 intervention schools had a trained Year 1 teacher (based on attending at least seven out of nine PD sessions) who remained teaching Year 1 for the majority of the year (2022–2023). (Four of the schools reported having a change in Year 1 teacher part-way through the year or of having teachers who job shared so the data for the teacher who taught for most of the year was used, or in the case of a job share, the teacher who attended the most PD sessions.) Around 31 schools fulfilled the criteria of having a trained Reception teacher who remained teaching Reception for the majority of the 2021–2022 academic year and of having a trained Year 1 teacher who remained teaching Year 1 for the majority of the 2022–2023 academic year.

However, we cannot guarantee that the classes taught by the trained Year 1 teacher contained all children involved in the evaluation. For example, it was possible that in schools with a multi-class Year 1 group, some of the children enrolled in Reception may have moved to a class taught by a non-participating teacher in Year 1. This is, in part, due to schools experiencing higher than usual staff turnover due to the COVID-19 pandemic. To track this, we asked schools at the start of the academic year 2022–2023 to indicate the name of the teacher each participating child in their school was due to be taught maths by, in Year 1 that year. These data were provided by 17 intervention schools (total *n*=296 pupils). These data indicate that, on average, 67.6% of randomised pupils per school were taught in Year 1 by a trained teacher (based on attending seven or more sessions). This figure was the same when considering a trained teacher to be one who attended at least five sessions. Eleven of the 17 schools (64.7%) indicated that at least 75% of pupils in the evaluation (Cohort 1 Reception children for whom a baseline BAS3 ENC assessment was conducted) moved to a Year 1 class being taught maths by a TEEMUP trained Year 1 teacher in the 2022–2023 academic year.

Table 17: Summary of compliance data for Cohort 1

Elements of compliance	Intervention schools (*n*=38)
Reception teacher attended at least seven out of nine PD sessions, *n* (%)	36 (94.7)
Reception teacher attended at least five out of nine PD sessions, *n* (%)	37 (97.4)
Number of PD sessions attended by Reception teacher	
Mean (SD), minimum-maximum	8.2 (1.6), 0–9
Of which, mode of attendance:	
Attended face-to-face, mean (SD)	6.8 (2.2)
Attended a trainer-led online training workshop, mean (SD)	0.9 (1.4)
Self-reported independent catch-up, mean (SD)	0.5 (1.0)
Year 1 teacher attended at least seven out of nine PD sessions, *n* (%)	33 (86.8)
Year 1 teacher attended at least five out of nine PD sessions, *n* (%)	33 (86.8)
Number of PD sessions attended by Year 1 teacher[a]	

Elements of compliance	Intervention schools (*n*=38)
Mean (SD), minimum-maximum	7.2 (2.5), 0–9
Of which, mode of attendance:	
Attended face-to-face, mean (SD)	5.2 (3.5)
Attended a trainer-led online training workshop, mean (SD)	0.7 (1.6)
Attended a trainer-led Zoom 'new teacher' session in Autumn Term 2022, mean (SD)	0.9 (2.2)
Self-reported independent catch-up, mean (SD)	0.4 (1.3)
Number of mentor/coach visits hosted by Reception teacher	
Mean (SD), minimum-maximum	3.1 (1.4), 0–4
Of which, nature of visit:	
Face-to-face, mean (SD)	2.4 (1.1)
Online, mean (SD)	0.7 (0.5)
Percentage of mentor/coach visits prepared for	
Mean (SD), minimum-maximum	85.0 (22.8), 25–100
Mentor/coach visits and preparedness (Reception)	
Good[b]	31 (81.6)
Minimal[c]	4 (10.5)
Non-compliant	3 (7.9)
Number of mentor/coach visits hosted by Year 1 teacher[a]	
Mean (SD), minimum-maximum	3.2 (1.2), 0–4
Of which, nature of visit:	
Face-to-face, mean (SD)	2.5 (1.1)
Online, mean (SD)	0.7 (0.6)
Percentage of mentor/coach visits prepared for	
Mean (SD), minimum-maximum	81.5 (24.5), 25–100
Mentor/coach visits and preparedness (Year 1)	
Good[b]	27 (71.1)
Minimal[c]	7 (18.4)
Non-compliant	4 (10.5)
Number of school logins to the online knowledge base over the course of the intervention period	
Mean (SD), minimum-maximum	40.3 (26.5), 10–120
Mentor-determined evidence of change (Reception)	
Excellent	24 (63.2)
Good	8 (21.1)
Minimal	3 (7.9)
Non-compliant	3 (7.9)
Mentor-determined evidence of change (Year 1)	
Excellent	22 (57.9)
Good	6 (15.8)
Minimal	6 (15.8)
Non-compliant	4 (10.5)
Overall classification of compliance for school	**N=47**
Good	22 (46.8)
Minimal	8 (17.0)
Non-compliant	17 (36.2)

[a] *Across 40 teachers/classes as two schools had two Year 1 classes taking part.*
[b] *Good=School hosted three face-to-face or online visits from a mentor/coach, at least two of which teachers were well prepared for.*
[c] *Minimal=School hosted two face-to-face or online visits from a mentor/coach, at least one of which teachers were well prepared for.*

The CACE estimate of the effect of the school being a 'good' complier on the pupils' maths attainment was a predicted increase of 1.37 points in BAS3 ENC post-test score (95% CI: -1.13 to 3.86, p = 0.28; effect size 0.12, 95% CI: -0.10 to 0.35, equating to two months' additional progress in the intervention group). The partial R^2 from the first stage of the CACE analysis was 0.35 and the F-statistic was F(1, 88 = 46.7, p<0.001); these indicate moderate correlation between the instrumental variable (random allocation) and the endogenous variable, and that the inference of the CACE estimate is reliable.

The CACE estimate of the effect of the school having at least minimal compliance on the pupils' maths attainment was a predicted increase of 1.02 points in BAS3 ENC post-test score (95% CI: -0.84 to 2.88, p = 0.28; effect size 0.09, 95% CI: -0.08 to 0.26, equating to one month's additional progress in the intervention group). The partial R^2 from the first stage of the CACE analysis was 0.52 and the F-statistic was F(1, 88 = 93.8, p<0.001); these indicate reasonable correlation between the instrumental variable (random allocation) and the endogenous variable, and that the inference of the CACE estimate is reliable.

In post-hoc sensitivity analyses, these CACE analyses were repeated assuming that schools that did not provide Year 1 teacher data were 'non-compliant'. Eight schools previously rated 'good' and three previously rated 'minimal' moved to 'non-compliant'—total eight 'good', three 'minimal', and 36 'non-compliant'. The CACE estimate of the effect of the school being a 'good' complier on the pupils' maths attainment was a predicted increase of 3.90 points in BAS3 ENC post-test score (95% CI: -3.52 to 11.32, p = 0.30; effect size 0.35, 95% CI: -0.32 to 1.03, equating to four months' additional progress in the intervention group). The partial R^2 from the first stage of the CACE analysis was 0.10 and the F-statistic was F(1, 88 = 10.0, p = 0.002); these indicate a weak correlation between the instrumental variable (random allocation) and the endogenous variable, and that the inference of the CACE estimate may not be reliable (the F-statistic should exceed ten for inference based on the 2SLS estimator to be reliable when there is one endogenous regressor, as in this case).

The CACE estimate of the effect of the school having at least minimal compliance on the pupils' maths attainment was a predicted increase of 2.85 points in BAS3 ENC post-test score (95% CI: -2.51 to 8.22, p = 0.30; effect size 0.26, 95% CI: -0.23 to 0.75, equating to three months' additional progress in the intervention group). The partial R^2 from the first stage of the CACE analysis was 0.15 and the F-statistic was F(1, 88 = 15.7, p<0.001); these indicate weak correlation between the instrumental variable (random allocation) and the endogenous variable, and that the inference of the CACE estimate is reliable.

Additional analyses and robustness checks

The primary analysis model was rerun with the additional covariate of whether the school was partaking in the NCETM programme. Fifty-three schools (57.0%; intervention, *n*=25/47, 53.2%; control, *n*=28/49, 60.9%) reported taking part in this programme during the trial period. The adjusted mean difference in BAS3 ENC post-test score between the intervention and control groups was 0.75 (95% CI: -0.53 to 2.04, p = 0.25). The estimated Hedges' g effect size was -0.07 (95% CI: -0.05 to 0.19), which relates to one additional month's progress in the intervention group. The total variance used to calculate the effect size was 121.26; the sum of 111.21 (random variation between pupils, within-cluster variance) and 10.05 (heterogeneity between schools, between-cluster variance). Since the effect size in this analysis is unchanged from the primary results, we conclude that any participation in the NCETM did not impact the TEEMUP results.

Missing data analysis

A mixed effect logistic regression model suggested that a 'missing at random' assumption was reasonable, meaning that MI by chained equations was an appropriate approach.

Among the randomised population with baseline BAS3 ENC score, pupils eligible for FSM were almost twice as likely to have missing post-test BAS3 ENC data than non-FSM pupils (adjusted OR 1.77, 95% CI: 1.19 to 2.65, p = 0.005). Additionally, pupils with higher CSBQ sociability scores were less likely to have missing data (OR 0.69, 95% CI: 0.49 to 0.98, p = 0.04), and those with higher externalising scores and emotional scores were more likely to have missing data (OR 1.52, 95% CI: 1.02 to 2.27, p = 0.04 and 2.04, 95% CI: 1.35 to 3.09, p = 0.001, respectively).

To investigate the impact of missing data, the primary analysis was repeated using MI by chained equations. The adjusted mean difference in BAS3 ENC score following MI was 0.50 (95% CI: -0.74 to 1.73, p = 0.43) and the Hedges' g effect size was 0.05 (95% CI: -0.07 to 0.16), which is similar to the ITT estimate.

Secondary analysis

CSBQ

Pupil characteristics for Cohort 1 pupils for whom valid baseline CSBQ data were collected are presented in Table 18 by group (intervention, *n*=975; control, *n*=979). The (absolute) effect size for the difference in baseline CSBQ scores between the groups ranged from 0.06 to 0.18.

Table 18: Baseline pupil characteristics and CSBQ scores for Cohort 1 pupils for whom baseline CSBQ data are available

Cohort 1	Intervention (*n*=975)	Control (*n*=979)	
Pupil level (categorical)	Count (%)	Count (%)	
Gender			
Male	507 (52.0)	477 (48.7)	
Female	468 (48.0)	502 (51.3)	
Eligible for FSM	134 (13.7)	168 (17.2)	
EAL	133 (13.6)	180 (18.4)	
Pupil level (continuous)	Mean (SD)	Mean (SD)	Effect size
Age (years)	4.3 (0.4)	4.3 (0.4)	—
CSBQ Sociability	3.8 (0.8)	3.6 (0.8)	0.17 (0.08, 0.26)
CSBQ Externalising problems	1.5 (0.8)	1.6 (0.9)	-0.06 (-0.15, 0.02)
CSBQ Internalising problems	1.5 (0.6)	1.6 (0.7)	-0.12 (-0.21, -0.03)
CSBQ Prosocial behaviour	3.8 (0.9)	3.6 (0.9)	0.18 (0.09, 0.27)
CSBQ Behavioural self-regulation	3.9 (0.9)	3.8 (0.9)	0.15 (0.06, 0.23)
CSBQ Cognitive self-regulation	3.4 (1.0)	3.3 (0.9)	0.11 (0.02, 0.20)
CSBQ Emotional self-regulation	4.0 (0.8)	3.9 (0.8)	0.17 (0.08, 0.26)
CSBQ Composite self-regulation	3.8 (0.8)	3.7 (0.8)	0.16 (0.07, 0.25)

The correlation between the baseline and follow-up CSBQ scores ranged from 0.56 to 0.73 at the end of Reception, but were slightly lower at the end of Year 1 (range 0.30 to 0.57) (Table 19).

Table 19: Correlation between baseline and post-test CSBQ scores at the end of Reception and Year 1 (Cohort 1)

Correlation between pre- and post-test scores (95% CI)	End of Reception	End of Year 1
CSBQ Sociability	0.70 (0.67, 0.73)	0.50 (0.46, 0.54)
CSBQ Externalising	0.71 (0.68, 0.74)	0.50 (0.46, 0.54)
CSBQ Internalising	0.56 (0.52, 0.60)	0.30 (0.26, 0.35)
CSBQ Prosocial	0.69 (0.66, 0.72)	0.52 (0.48, 0.55)
CSBQ Behavioural	0.73 (0.70, 0.76)	0.59 (0.55, 0.62)
CSBQ Cognitive	0.64 (0.60, 0.67)	0.49 (0.45, 0.53)
CSBQ Emotional	0.69 (0.66, 0.72)	0.48 (0.44, 0.52)
CSBQ Self-regulation	0.73 (0.70, 0.75)	0.57 (0.53, 0.60)

Mean CSBQ scores are presented by group and time point in Table 20. The unadjusted and adjusted mean differences are presented in Appendix D Table 1. The estimated Hedges' g effect sizes favoured the intervention group at the end of Reception and ranged in (absolute) magnitude from 0.13 to 0.44 (approximately two to five months' additional progress). All differences at this time point were statistically significant at the 5% level except for the Prosocial scale. However, at the end of Year 1, the benefits largely reversed to favour the control group, except for the Externalising, Internalising, and Emotional subscales, where the effect size favouring the intervention group ranged in (absolute) magnitude from 0.05 to 0.15 (approximately one to two months' additional progress). Only the differences for the Externalising and Prosocial scales were statistically significant at this time point. The total variances used to calculate the effect sizes are reported in Appendix D Table 1.

At the post-test CSBQ time point, the assessment was completed for the pupil by their current teacher in 95% of cases. The remaining cases were mostly completed by the assistant headteacher of the school, a teaching assistant, or the teacher of a parallel form.

Table 20: Cohort 1 CSBQ secondary outcomes

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	*n* (missing)	Mean (95% CI)	*n* (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	P-value	Estimated months' progress
End of Reception								
CSBQ Sociability	593 (382)	4.2 (4.1, 4.2)	563 (416)	3.9 (3.9, 4.0)	1,138 (593; 545)	0.20 (0.06, 0.34)	0.004	3
CSBQ Externalising	593 (382)	1.4 (1.4, 1.5)	563 (416)	1.6 (1.6, 1.7)	1,138 (593; 545)	-0.29 (-0.42, -0.15)	<0.001	4
CSBQ Internalising	592 (383)	1.3 (1.3, 1.4)	563 (416)	1.6 (1.5, 1.7)	1,137 (592; 545)	-0.44 (-0.63, -0.25)	<0.001	5
CSBQ Prosocial	592 (383)	4.2 (4.1, 4.3)	563 (416)	3.9 (3.9, 4.0)	1,137 (592; 545)	0.13 (-0.01, 0.28)	0.07	2
CSBQ Behavioural	593 (382)	4.3 (4.2, 4.4)	563 (416)	4.0 (3.9, 4.1)	1,138 (593; 545)	0.23 (0.09, 0.38)	0.002	3
CSBQ Cognitive	593 (382)	3.9 (3.8, 3.9)	563 (416)	3.6 (3.5, 3.6)	1,138 (593; 545)	0.23 (0.07, 0.38)	0.004	3
CSBQ Emotional	593 (382)	4.2 (4.2, 4.3)	563 (416)	4.0 (3.9, 4.0)	1,138 (593; 545)	0.20 (0.05, 0.35)	0.01	3
CSBQ Self-regulation	593 (382)	4.1 (4.1, 4.2)	563 (416)	3.8 (3.8, 3.9)	1,138 (593; 545)	0.25 (0.09, 0.40)	0.002	3
End of Year 1								
CSBQ Sociability	709 (266)	4.0 (3.9, 4.1)	659 (320)	3.9 (3.9, 4.0)	1,347 (709; 638)	-0.08 (-0.23, 0.06)	0.27	-1
CSBQ Externalising	709 (266)	1.5 (1.4, 1.6)	659 (320)	1.6 (1.6, 1.7)	1,347 (709; 638)	-0.15 (-0.29, -0.01)	0.03	2
CSBQ Internalising	709 (266)	1.6 (1.5, 1.6)	659 (320)	1.7 (1.7, 1.8)	1,347 (709; 638)	-0.15 (-0.33, 0.02)	0.08	2
CSBQ Prosocial	709 (266)	4.0 (3.9, 4.1)	659 (320)	4.0 (3.9, 4.0)	1,347 (709; 638)	-0.15 (-0.29, -0.00)	0.04	-2
CSBQ Behavioural	709 (266)	4.0 (3.9, 4.1)	659 (320)	4.0 (3.9, 4.0)	1,347 (709; 638)	-0.07 (-0.21, 0.07)	0.34	-1

CSBQ Cognitive	709 (266)	3.6 (3.5, 3.6)	659 (320)	3.5 (3.4, 3.6)	1,347 (709; 638)	-0.05 (-0.20, 0.10)	0.52	-1
CSBQ Emotional	709 (266)	4.1 (4.0, 4.2)	659 (320)	4.0 (3.9, 4.0)	1,347 (709; 638)	0.05 (-0.11, 0.20)	0.55	1
CSBQ Self-regulation	709 (266)	3.9 (3.8, 3.9)	659 (320)	3.8 (3.8, 3.9)	1,347 (709; 638)	-0.04 (-0.19, 0.11)	0.63	0

The ICC associated with school for the baseline CSBQ scores ranged from 0.08 to 0.17 (Table 21). The ICC associated with school from the adjusted model at the post-test at the end of Year 1 increased (range 0.11 to 0.23).

Table 21: ICC associated with school for the CSBQ secondary outcomes for Cohort 1 at pre- and post-tests

ICC (95% CI)	Pre-test	End of Reception		End of Year 1	
CSBQ subscale		Empty model	Adjusted model	Empty model	Adjusted model
CSBQ Sociability	0.12 (0.09, 0.17)	0.21 (0.15, 0.30)	0.15 (0.10, 0.22)	0.11 (0.07, 0.17)	0.13 (0.08, 0.18)
CSBQ Externalising	0.09 (0.06, 0.13)	0.15 (0.10, 0.24)	0.15 (0.09, 0.22)	0.08 (0.04, 0.13)	0.11 (0.07, 0.16)
CSBQ Internalising	0.17 (0.11, 0.25)	0.27 (0.20, 0.37)	0.14 (0.08, 0.24)	0.22 (0.16, 0.30)	0.21 (0.15, 0.29)
CSBQ Prosocial	0.12 (0.08, 0.17)	0.17 (0.12, 0.24)	0.12 (0.08, 0.19)	0.14 (0.10, 0.20)	0.19 (0.14, 0.25)
CSBQ Behavioural	0.09 (0.06, 0.13)	0.14 (0.10, 0.20)	0.14 (0.09, 0.20)	0.12 (0.08, 0.17)	0.21 (0.16, 0.27)
CSBQ Cognitive	0.08 (0.05, 0.12)	0.19 (0.14, 0.26)	0.17 (0.10, 0.26)	0.08 (0.05, 0.12)	0.15 (0.11, 0.20)
CSBQ Emotional	0.09 (0.07, 0.12)	0.17 (0.12, 0.24)	0.12 (0.07, 0.19)	0.13 (0.08, 0.18)	0.17 (0.12, 0.23)
CSBQ Self-regulation	0.08 (0.06, 0.12)	0.19 (0.14, 0.26)	0.18 (0.12, 0.25)	0.12 (0.08, 0.18)	0.23 (0.17, 0.29)

EYFSP

In total, valid NPD data were available for 1,557 randomised pupils (99.4%; intervention *n* = 776/780, 99.5%; control *n* = 781/786, 99.4%), from 93 schools (47 intervention, 46 control). For the EYFSP ELG average point score, a mean of 1.88 (95% CI: 1.87 to 1.90) was observed in the intervention arm and 1.88 (95% CI: 1.87 to 1.90) in the control arm. The unadjusted mean difference (95% CI) is 0.00 (-0.02 to 0.02) (Table 22). The correlation between the baseline BAS3 ENC and EYFSP ELG average point outcome score was 0.37 (95% CI: 0.33 to 0.42). The adjusted mean difference in post-test score between the intervention and control groups was 0.0 (95% CI: -0.03 to 0.03, p = 0.99, Table 22).

The estimated Hedges' g effect size was 0.00 (95% CI: -0.13 to 0.13), which does not relate to any additional months' progress in the intervention group (Table 22). The total variance used to calculate the effect size was 0.05 (Appendix D Table 1); the sum of 0.05 (random variation between pupils, within-cluster variance) and 0.00 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.04 (95% CI: 0.02 to 0.09). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.06 (0.03 to 0.10).

The estimated Hedges' effect sizes between the intervention and control groups for the maths, self-regulation, and PSED ELG outcomes were 0.05 (95% CI: -0.16 to 0.26), -0.01 (95% CI: -0.28 to 0.26) and 0.05 (95% CI: -0.20 to 0.30), respectively, equating to zero to one month's additional progress in the intervention group (Table 22).

Similar proportions of pupils in the intervention and control groups achieved a Good Level of Development overall on the EYFSP (68.4% and 69.1%, respectively). The estimated Hedges' effect size was -0.01 (95% CI: -0.22 to 0.19), which does not relate to any additional months' progress in the intervention group (Table 22).

Table 22: EYFSP secondary outcomes for Cohort 1

EYFSP secondary outcomes	Intervention group (*n*=776) N (%)	Control group (*n*=781) N (%)	Correlation with baseline BAS3 ENC	Unadjusted percentage point difference (95% CI)	Adjusted percentage point difference (95% CI)	Adjusted OR (95% CI)	Hedges' g (95% CI)	P-value
Maths ELG	624 (80.4)	626 (80.2)	0.37 (0.32, 0.41)	0.26 (-3.69, 4.21)	1.01 (-3.21, 5.23)	1.09 (0.77, 1.54)	0.05 (-0.16, 0.26)	0.64
Self-regulation ELG	697 (89.8)	703 (90.0)	0.17 (0.12, 0.22)	-0.19 (-3.18, 2.80)	-0.14 (-3.88, 3.59)	0.98 (0.63, 1.54)	-0.01 (-0.28, 0.26)	0.94
PSED ELG	687 (88.5)	686 (87.8)	0.19 (0.14, 0.24)	0.69 (-2.51, 3.90)	0.81 (-3.13, 4.74)	1.09 (0.72, 1.65)	0.05 (-0.20, 0.30)	0.69
Good Level of Development	531 (68.4)	540 (69.1)	0.36 (0.31, 0.40)	-0.71 (-5.32, 3.89)	-0.30 (-5.97, 5.36)	0.98 (0.70, 1.38)	-0.01 (-0.22, 0.19)	0.92
	Mean (95% CI)	Mean (95% CI)	Correlation with baseline BAS3 ENC	Unadjusted mean difference (95% CI)	Adjusted mean difference (95% CI)	—	Hedges' g (95% CI)	P-value
ELG mean	1.88 (1.87, 1.90)	1.88 (1.87, 1.90)	0.37 (0.33, 0.42)	0.00 (-0.02, 0.02)	0.00 (-0.03, 0.03)	—	0.00 (-0.13, 0.13)	0.99

In sensitivity analyses, the analysis models for EYFSP Self-regulation and PSED ELG outcomes were repeated adjusting for baseline CSBQ scores for these domains. As expected, the correlation between the self-regulation ELG and the self-regulation component of the CSBQ was higher (at 0.46, Table 23) than for the BAS3 ENC baseline score (0.20, Table 22). In the adjusted analysed, the magnitude of the Hedges' g effect size was substantially increased (relative to when the BAS3 ENC was adjusted for) and favoured the control group (-0.33, 95% CI: -0.69 to 0.05, p = 0.08). Similarly, the PSED ELG was more highly correlated with each baseline CSBQ subscale score (absolute range 0.28 to 0.46, Table 23) than it was with the BAS3 ENC score (0.22, Table 22), and in the adjusted analyses, the Hedges' g effect size favoured the control group, though none of the differences were statistically significant.

Table 23: Sensitivity analyses for EYFSP outcomes for Cohort 1

Sensitivity EYFSP analyses	Correlation (95% CI)	Adjusted percentage point difference (95% CI)	Adjusted OR (95% CI)	Hedges' g (95% CI)	P-value
Self-regulation ELG, adjusting for CSBQ self-regulation	0.46 (0.42, 0.50)	-3.12 (-6.66, 0.42)	0.58 (0.32, 1.08)	-0.33 (-0.69, 0.05)	0.08
PSED ELG, adjusting for:					
CSBQ Sociability	0.28 (0.23, 0.33)	-1.35 (-5.17, 2.47)	0.86 (0.56, 1.32)	-0.09 (-0.35, 0.17)	0.49
CSBQ Externalising	-0.35 (-0.40, -0.31)	-1.14 (-4.81, 2.54)	0.88 (0.57, 1.34)	-0.08 (-0.34, 0.18)	0.54
CSBQ Internalising	-0.34 (-0.39, -0.30)	-1.20 (-5.03, 2.64)	0.87 (0.56, 1.36)	-0.08 (-0.35, 0.19)	0.54
CSBQ Prosocial	0.41 (0.37, 0.45)	-2.15 (-5.95, 1.64)	0.75 (0.45, 1.25)	-0.17 (-0.48, 0.13)	0.27
CSBQ Behavioural	0.46 (0.42, 0.50)	-1.89 (-5.46, 1.68)	0.77 (0.47, 1.26)	-0.16 (-0.46, 0.14)	0.30
CSBQ Cognitive	0.37 (0.33, 0.41)	-0.65 (-4.60, 3.31)	0.92 (0.56, 1.52)	-0.05 (-0.35, 0.25)	0.75
CSBQ Emotional	0.40	-2.36	0.74	-0.18	0.20

	(0.36, 0.45)	(-5.94, 1.22)	(0.48, 1.17)	(-0.45, 0.09)	

Cohort 2

BAS3 ENC analysis

In total, a valid post-test BAS3 ENC score was available for 1,001 pupils in Cohort 2 (intervention, n = 525; control, n = 476), from 73 schools (38 intervention, 35 control). A mean of 45.6 (95% CI: 44.8 to 46.4) was observed in the intervention arm and 45.7 (95% CI: 44.9 to 46.5) in the control arm. The unadjusted mean difference (95% CI) is -0.08 (-1.19 to 1.03). The mean school-average baseline BAS3 ENC score from Cohort 1 was 44.7 (SD 2.7) in the intervention group and 44.7 (SD 2.7) in the control group. Histograms of the baseline and post-test scores are provided (Figure 5). The correlation between the baseline and post-test scores was 0.11 (95% CI 0.05 to 0.17). Around 983pupils (intervention, *n*=510; control, *n*=473) were included in the Cohort 2 BAS3 ENC analysis as 18 were missing EAL covariate data. As a check of the analysis model assumptions, the normality of the standardised residuals was checked using a QQ plot and demonstrated a good fit to the normal distribution (Figure 6). The adjusted mean difference in post-test score between the intervention and control groups was 0.08 (95% CI: -1.28 to 1.45, p = 0.90, Appendix D Table 2). The estimated Hedges' g effect size was 0.01 (95% CI: -0.14 to 0.16), which does not relate to any additional months' progress. The total variance used to calculate the effect size was 79.72; the sum of 74.15 (random variation between pupils, within-cluster variance) and 5.57 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.04 (95% CI: 0.02 to 0.10). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.07 (0.04 to 0.12).

Omitting the school-level mean BAS3 ENC from Cohort 1 as a covariate

The adjusted mean difference in post-test score between the intervention and control groups for the model excluding baseline BAS3 ENC was 0.07 (95% CI: -1.33 to 1.48, p = 0.92). The estimated Hedges' g effect size was 0.01 (95% CI: -0.15 to 0.17), which does not relate to any additional months' progress in the intervention group (Table 28) and is very similar to the estimate from the analysis adjusting for school-level mean BAS3 ENC from Cohort 1 as a covariate. The ICC associated with school from the adjusted model was 0.04 (95% CI: 0.02 to 0.10).

Figure 5: Histograms of school-level mean baseline BAS3 ENC scores from Cohort 1 and post-test BAS3 ENC scores for Cohort 2 (scores have been aggregated to present the mean score for every ten pupils to prevent statistical disclosure)

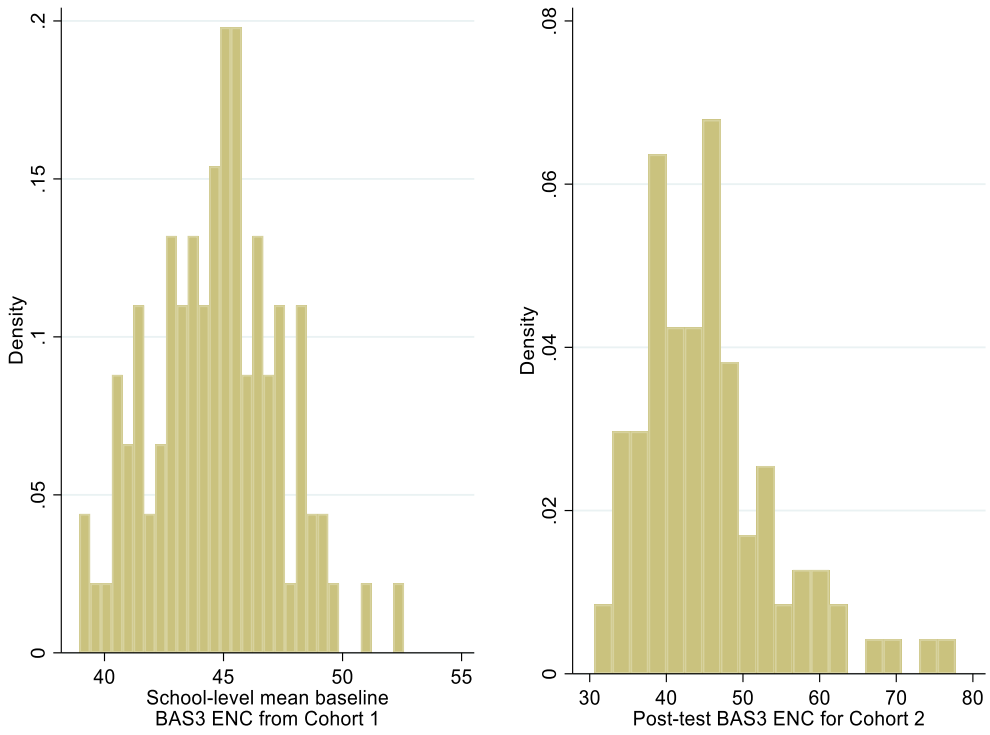
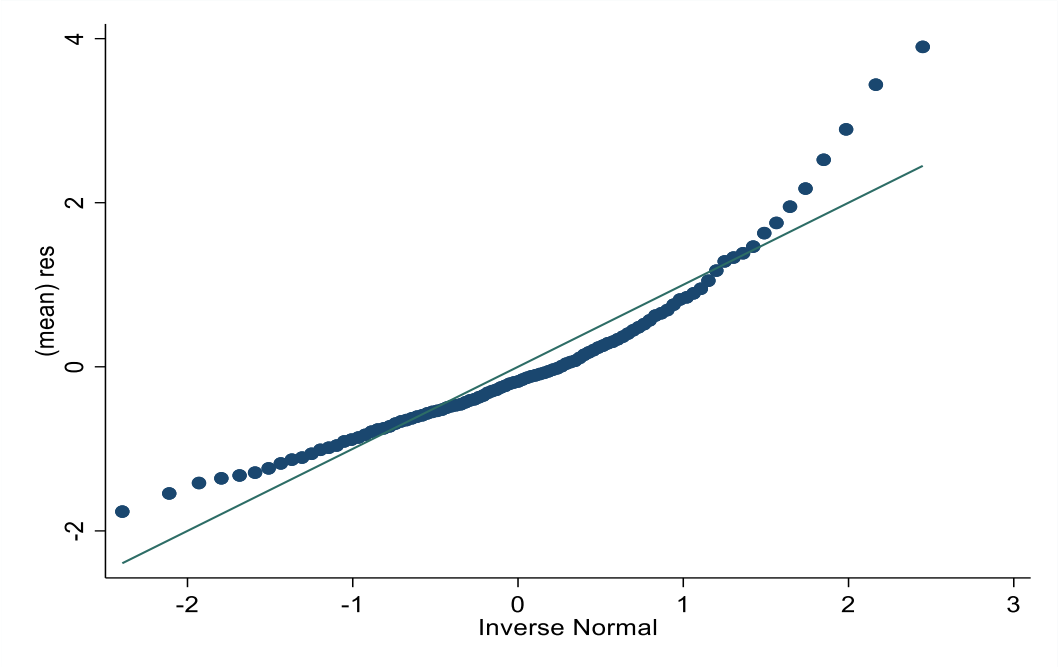


Figure 6: QQ plot of the standardised residuals from the BAS3 ENC analysis model for Cohort 2 to assess the normality assumption (residuals have been aggregated to present the mean residual for every ten pupils to prevent statistical disclosure)



Subgroup analyses

Summary statistics for Cohort 2 for the post-test BAS3 ENC score are presented by gender, FSM eligibility, and whether schools were taking part in the NCETM Maths Hubs Programme (Table 24). These summaries indicate that, in general, pupils eligible for FSM performed slightly worse on the post-test than those not eligible for FSM. Scores were reasonably similar for male and female pupils, and for pupils stratified by whether their school participated in the NCETM Maths Hubs Programme.

In a series of adjusted regression analyses that included interaction effects, there was no evidence of an interaction between trial allocation and: i) FSM eligibility (-0.91, 95% CI: -3.22 to 1.40, p = 0.44); or ii) school participation in NCETM Maths Hubs Programme (-0.62, 95% CI: -4.35 to 3.11, p = 0.74). There was, however, evidence, of an interaction for gender (interaction term coefficient 2.21, 95% CI: 0.07 to 4.35, p = 0.04); the adjusted mean difference is positive for males (1.21, 95% CI: -0.60 to 3.02, p = 0.19; estimated Hedges' g 0.14, 95% CI: -0.07 to 0.34) but negative for females (-1.00, 95% CI: -2.66 to 0.66, p = 0.24; estimated Hedges' g -0.11, 95% CI: -0.30 to 0.07).

Within the sample with post-test BAS3 ENC scores, 177 pupils from 73 schools were eligible for FSM (median per school 2, range 0 to 7 in both groups). Within this restricted sample of pupils eligible for FSM, the correlation between the baseline and post-test scores was 0.09 (95% CI: -0.06 to 0.24). The adjusted mean difference in post-test score between the intervention and control groups within this sample was -0.41 (95% CI: -2.64 to 1.82, p = 0.72, Appendix D Table 2) and the estimated Hedges' g effect size was -0.06 (95% CI: -0.37 to 0.26), which relates to one fewer month's progress in the intervention group. The total variance used to calculate the effect size (from a model without covariates) was 49.60—the sum of 45.32 (random variation between children, within-cluster variance) and 4.28 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.03 (95% CI: 0.00 to 0.97). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.09 (0.01 to 0.54). As a sensitivity check, the effect size for the FSM-eligible subgroup from the interaction effects model was extracted and compared with the effect size derived from the restricted sample. The adjusted mean difference in post-test score between the intervention and control groups within this sample was -0.66 (95% CI: -3.00 to 1.68, p = 0.58) and the estimated Hedges' g effect size was -0.07 (95% CI: -0.34 to 0.19), which relates to one fewer month's progress in the intervention group.

Table 24: Subgroup summary scores for the post-test BAS3 ENC score

Outcome	Raw means			
	Intervention		Control	
	n	Mean (SD)	N	Mean (SD)
Gender				
Male	248	46.0 (10.2)	243	45.2 (9.6)
Female	277	45.2 (7.6)	233	46.1 (8.2)
FSM eligibility				
Yes	93	41.5 (6.9)	84	42.3 (7.3)
No	432	46.5 (9.1)	392	46.4 (9.1)
School's participation in NCETM Maths Hubs Programme				
Yes	279	45.3 (8.7)	303	45.5 (8.8)
No	246	45.9 (9.2)	173	46.0 (9.2)

Analysis in the presence of non-compliance

Overall, including the nine schools that withdrew from the intervention, 27 of the 47 intervention schools (57.5%) were classed as having 'good' compliance for Cohort 2, four (8.5%) as 'minimal', and 16 (34.0%) as non-compliant. All 46 control schools were classed as non-compliant (Table 25).

30 of the 38 intervention schools had a trained Reception teacher (based on attending at least seven out of nine PD sessions) who remained teaching Reception for the majority of the year (2022–2023) (Table 25). Six of the schools were reported as having a change in Reception teacher part-way through Reception so the data for the teacher who taught for most of the year was used. Three-quarters (*n*=28, 73.7%) of the schools retained a trained Reception teacher from 2021–2022 to teach for the majority of the year in Reception in 2022–2023.

Table 25: Summary of compliance data for Cohort 2

Elements of compliance	Intervention schools (*n*=38)
Attended at least seven out of nine PD sessions, *n* (%)	30 (78.9)
Attended at least five out of nine PD sessions, *n* (%)	31 (81.6)
Number of PD sessions attended by Reception teacher	
Mean (SD), minimum-maximum	6.9 (3.3), 0-9
Of which, mode of attendance:	
Attended face-to-face, mean (SD)	5.5 (3.5)
Attended a trainer-led online training workshop, mean (SD)	0.5 (0.7)
Attended a trainer-led Zoom 'new teacher' session in Autumn Term 2022, mean (SD)	0.5 (1.9)
Self-reported independent catch-up, mean (SD)	0.4 (1.1)
Number of mentor/coach visits hosted by Reception teacher	
Mean (SD), minimum-maximum	3.1 (1.3), 0-4
Of which, nature of visit:	
Face-to-face, mean (SD)	2.4 (1.1)
Online, mean (SD)	0.7 (0.5)
Percentage of mentor/coach visits prepared for	
Mean (SD), minimum-maximum	79.2 (29.5), 0-100
Mentor/coach visits and preparedness (Reception)	
Good[a]	30 (78.9)
Minimal[b]	3 (7.9)
Non-compliant	5 (13.2)
Number of school logins to the online knowledge base over the course of the intervention period	
Mean (SD), minimum-maximum	39.3 (25.8), 10-120
Mentor-determined evidence of change (Reception)	
Excellent	25 (65.8)
Good	5 (13.2)

Elements of compliance	Intervention schools (*n*=38)
Minimal	3 (7.9)
Non-compliant	5 (13.2)
Overall classification of compliance for school	**N=47**
Good	27 (57.5)
Minimal	4 (8.5)
Non-compliant	16 (34.0)

[a]*Good=School hosted three face-to-face or online visits from a mentor/coach, at least two of which teachers were well prepared for.*
[b]*Minimal=School hosted two face-to-face or online visits from a mentor/coach, at least one of which teachers were well prepared for.*

The CACE estimate of the effect of the school being a 'good' complier on the pupils' maths attainment was a predicted increase of 0.14 points in BAS3 ENC post-test score (95% CI: -1.77 to 2.04, p = 0.89; effect size 0.02, 95% CI: -0.20 to 0.23, equating to zero month's additional progress in the intervention group). The partial R^2 from the first stage of the CACE analysis was 0.56 and the F-statistic was F(1, 71 = 95.1, p<0.001); these indicate reasonable correlation between the instrumental variable (random allocation) and the endogenous variable, and that the inference of the CACE estimate is reliable.

The CACE estimate of the effect of the school having at least minimal compliance on the pupils' maths attainment was a predicted increase of 0.12 points in BAS3 ENC post-test score (95% CI: -1.54 to 1.77, p = 0.89; effect size 0.01, 95% CI: -0.17 to 0.20, equating to zero month's additional progress in the intervention group). The partial R^2 from the first stage of the CACE analysis was 0.71 and the F-statistic was F(1, 71 = 181.47, p<0.001); these indicate good correlation between the instrumental variable (random allocation), and the endogenous variable and that the inference of the CACE estimate is reliable.

Additional analyses and robustness checks

The BAS3 ENC analysis model for Cohort 2 was rerun with the additional covariate of whether the school was partaking in the NCETM programme. The adjusted mean difference in post-test score between the intervention and control groups was 0.03 (95% CI: -1.37 to 1.43, p = 0.97). The estimated Hedges' g effect size was 0.00 (95% CI: -0.15 to 0.16), which does not relate to any additional months' progress in the intervention group. The total variance used to calculate the effect size (from a model without covariates) was 79.72—the sum of 74.15 (random variation between children, within-cluster variance) and 5.57 (heterogeneity between schools, between-cluster variance).

CSBQ

Pupil characteristics for Cohort 2 pupils for whom valid baseline CSBQ data were collected are presented in Table 26 by group (intervention, *n*=484; control, *n*=485). The (absolute) effect size for the difference in baseline CSBQ scores between the groups ranged from 0.22 to 0.35.

Table 26: Baseline characteristics and CSBQ scores for Cohort 2 pupils for whom baseline CSBQ data are available

Cohort 2	Intervention (*n*=484)	Control (*n*=485)	
Pupil level (categorical)	Count (%)	Count (%)	
Gender			
Male	233 (48.1)	240 (49.5)	
Female	251 (51.9)	245 (50.5)	
Eligible for FSM	90 (18.6)	92 (19.0)	
EAL	82 (16.9)	70 (14.4)	
Pupil level (continuous)	Mean (SD)	Mean (SD)	Effect size
CSBQ Sociability	3.7 (0.8)	3.5 (0.9)	0.23 (0.10, 0.35)
CSBQ Externalising problems	1.5 (0.8)	1.7 (0.9)	-0.24 (-0.36, -0.11)

CSBQ Internalising problems	1.6 (0.7)	1.8 (0.7)	-0.23 (-0.35, -0.10)
CSBQ Prosocial behaviour	3.8 (0.9)	3.6 (0.9)	0.24 (0.11, 0.37)
CSBQ Behavioural self-regulation	3.9 (0.9)	3.7 (1.0)	0.26 (0.13, 0.38)
CSBQ Cognitive self-regulation	3.4 (1.0)	3.2 (1.0)	0.22 (0.09, 0.34)
CSBQ Emotional self-regulation	4.1 (0.8)	3.8 (0.9)	0.35 (0.22, 0.47)
CSBQ Composite self-regulation	3.8 (0.8)	3.6 (0.8)	0.31 (0.19, 0.44)

The correlation between the baseline and follow-up CSBQ scores ranged from 0.57 to 0.79 (Table 27).

Table 27: Correlation between baseline and post-test CSBQ scores at the end of Reception (Cohort 2)

Correlation between pre- and post-test scores (95% CI)	Post-test
CSBQ Sociability	0.70 (0.66, 0.73)
CSBQ Externalising	0.74 (0.70, 0.77)
CSBQ Internalising	0.57 (0.52, 0.61)
CSBQ Prosocial	0.72 (0.68, 0.75)
CSBQ Behavioural	0.77 (0.74, 0.79)
CSBQ Cognitive	0.71 (0.68, 0.74)
CSBQ Emotional	0.74 (0.71, 0.77)
CSBQ Self-regulation	0.79 (0.76, 0.81)

Mean CSBQ scores are presented by group and time point in Table 28. The unadjusted and adjusted mean differences are presented in Appendix D Table 2. The estimated Hedges' g effect sizes favoured the intervention group at the end of Reception and ranged in (absolute) magnitude from 0.10 to 0.39 (approximately two to four months' additional progress). All differences at this time point were statistically significant at the 5% level except for the Externalising scale. The total variances used to calculate the effect sizes are reported in Appendix D Table 2.

At the post-test CSBQ time point, the assessment was completed for the child by their current teacher in 97% of cases. The remaining cases were mostly completed by a teaching assistant.

Table 28: Cohort 2 secondary outcomes

Outcome	Unadjusted means				Effect size			
	Intervention group		Control group					
	n (missing)	Mean (95% CI)	*n* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges' g (95% CI)	P-value	Estimated months' progress
BAS3 ENC (adjusting for prior attainment)	525 (0)	45.6 (44.8, 46.4)	476 (0)	45.7 (44.9, 46.5)	983 (510; 473)	0.01 (-0.14, 0.16)	0.90	0
BAS3 ENC (not adjusting for prior attainment)	525 (0)	45.6 (44.8, 46.4)	476 (0)	45.7 (44.9, 46.5)	983 (510; 473)	0.01 (-0.15, 0.17)	0.92	0
	N=484		N=485					
CSBQ Sociability	430 (54)	4.1 (4.1, 4.2)	436 (49)	3.7 (3.6, 3.8)	863 (430; 433)	0.29 (0.13, 0.45)	0.00	4
CSBQ Externalising	430 (54)	1.4 (1.4, 1.5)	436 (49)	1.7 (1.6, 1.7)	863 (430; 433)	-0.10 (-0.28, 0.08)	0.27	2

CSBQ Internalising	430 (54)	1.4 (1.4, 1.5)	436 (49)	1.8 (1.7, 1.8)	863 (430; 433)	-0.34 (-0.52, -0.15)	0.00	4
CSBQ Prosocial	430 (54)	4.2 (4.1, 4.2)	436 (49)	3.8 (3.7, 3.9)	863 (430; 433)	0.18 (0.01, 0.36)	0.04	2
CSBQ Behavioural	430 (54)	4.2 (4.1, 4.3)	436 (49)	3.8 (3.7, 3.9)	863 (430; 433)	0.18 (0.02, 0.34)	0.02	2
CSBQ Cognitive	430 (54)	3.8 (3.7, 3.9)	436 (49)	3.4 (3.3, 3.5)	863 (430; 433)	0.23 (0.08, 0.36)	0.00	3
CSBQ Emotional	430 (54)	4.2 (4.1, 4.3)	436 (49)	3.8 (3.7, 3.9)	863 (430; 433)	0.18 (0.01, 0.34)	0.03	2
CSBQ Self-regulation	430 (54)	4.1 (4.0, 4.1)	436 (49)	3.7 (3.6, 3.8)	863 (430; 433)	0.21 (0.05, 0.37)	0.01	3

The ICC associated with school for the baseline CSBQ scores ranged from 0.08 to 0.20. The ICC associated with school from the adjusted model at the post-test at the end of Reception ranged from 0.09 to 0.16) (Table 29).

Table 29: ICC associated with school for the CSBQ secondary outcomes for Cohort 2 at pre- and post-tests

ICC (95% CI)	Pre-test	End of Reception	
CSBQ subscale		Empty model	Adjusted model
CSBQ Sociability	0.13 (0.08, 0.21)	0.17 (0.12, 0.23)	0.12 (0.07, 0.20)
CSBQ Externalising	0.09 (0.05, 0.15)	0.10 (0.06, 0.18)	0.13 (0.07, 0.21)
CSBQ Internalising	0.20 (0.14, 0.27)	0.22 (0.15, 0.30)	0.12 (0.07, 0.20)
CSBQ Prosocial	0.10 (0.05, 0.20)	0.09 (0.05, 0.15)	0.15 (0.10, 0.23)
CSBQ Behavioural	0.08 (0.04, 0.16)	0.06 (0.03, 0.12)	0.15 (0.09, 0.24)
CSBQ Cognitive	0.11 (0.06, 0.20)	0.10 (0.06, 0.15)	0.09 (0.04, 0.17)
CSBQ Emotional	0.13 (0.08, 0.22)	0.14 (0.08, 0.23)	0.14 (0.08, 0.23)
CSBQ Self-regulation	0.10 (0.05, 0.20)	0.11 (0.06, 0.18)	0.16 (0.10, 0.25)

In pre-planned sensitivity analysed, the CSBQ models were rerun adjusting additionally for an indicator for whether the baseline CSBQ data were collected before or after Christmas (Table 30). Results were virtually identical to those presented in Table 28.

Table 30: Cohort 2 CSBQ outcomes adjusting for timing of baseline CSBQ data collection

Outcome	Effect size		
	Adjusted mean difference (95% CI)	Hedges' g (95% CI)	P-value
CSBQ Sociability	0.24 (0.11, 0.37)	0.29 (0.13, 0.45)	0.00
CSBQ Externalising	-0.09 (-0.22, 0.05)	-0.12 (-0.28, 0.06)	0.22
CSBQ Internalising	-0.23 (-0.36, -0.10)	-0.35 (-0.54, -0.15)	0.00
CSBQ Prosocial	0.17 (0.01, 0.32)	0.20 (0.01, 0.37)	0.03
CSBQ Behavioural	0.18 (0.03, 0.33)	0.19 (0.03, 0.35)	0.02
CSBQ Cognitive	0.21 (0.08, 0.35)	0.22 (0.08, 0.37)	0.00
CSBQ Emotional	0.16 (0.02, 0.31)	0.19 (0.02, 0.37)	0.03
CSBQ Self-regulation	0.17 (0.05, 0.30)	0.22 (0.06, 0.38)	0.01

EYFSP

Pupil details were sent to NPD for matching for 1,060 Cohort 2 pupils who had valid baseline BAS3 ENC post-test data and/or CSBQ data at baseline and post-test (intervention, *n*=549; control, *n*=511), and were matched for 1,059 pupils (intervention, *n*=549; control, *n*=510) of which 1,056 had valid EYFSP data (intervention, *n*=546; control, *n*=510) from 73 schools (38 intervention, 35 control).

For the EYFSP ELG average point score, a mean of 1.90 (95% CI: 1.88 to 1.91) was observed in the intervention arm and 1.87 (95% CI: 1.84 to 1.89) in the control arm. The unadjusted mean difference (95% CI) is 0.03 (0.00 to 0.06). The correlation between the baseline BAS3 ENC and EYFSP ELG average point outcome score was 0.11 (95% CI: 0.05 to 0.17). The adjusted mean difference in post-test score between the intervention and control groups was 0.03 (95% CI: 0.01 to 0.05, p = 0.02, Table 31). The estimated Hedges' g effect size was 0.12 (95% CI: 0.04 to 0.20), which relates to approximately two months' additional progress. The total variance used to calculate the effect size was 0.06 (Appendix D Table 2); the sum of 0.06 (random variation between pupils, within-cluster variance) and 0.00 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.00 (95% CI: 0.00 to 0.00). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.02 (0.00 to 0.07).

The estimated Hedges' effect sizes for the difference between the intervention and control groups for the Self-regulation and PSED ELG outcomes were 0.03 (95% CI: -0.23 to 0.30) and 0.03 (95% CI: -0.22 to 0.28), respectively, equating to zero month's additional progress in the intervention group. However, there was a statistically significant difference for the Maths ELG favouring the intervention group (Hedges' effect size 0.27, 95% CI: 0.07 to 0.48, p = 0.01), equating to an approximate four months' additional progress (Table 31).

Similar proportions of pupils in the intervention and control groups achieved a Good Level of Development overall on the EYFSP (72.0% and 69.6%, respectively). The estimated Hedges' effect size was 0.06 (95% CI: -0.13 to 0.25), which relates to approximately one month's additional progress in the intervention group (Table 31).

Results were virtually unchanged when the EYFSP analyses were repeated omitting the mean baseline BAS3 ENC score per school from Cohort 1 (Table 32).

Table 31: EYFSP secondary outcomes for Cohort 2

	Intervention group (*n*=776) N (%)	Control group (*n*=781) N (%)	Correlation with baseline BAS3 ENC	Unadjusted percentage point difference (95% CI)	Adjusted percentage point difference (95% CI)	Adjusted OR (95% CI)	Hedges' g (95% CI)	P-value
Maths ELG	463 (84.8)	400 (78.4)	0.09 (0.03, 0.15)	6.37 (1.70,11.04)	6.44 (1.72,11.15)	1.57 (1.13, 2.20)	0.27 (0.07, 0.48)	0.01
Self-regulation ELG	486 (89.0)	448 (87.8)	0.06 (0.00, 0.12)	1.17 (-2.70, 5.03)	0.60 (-3.59, 4.80)	1.06 (0.69, 1.64)	0.03 (-0.23, 0.30)	0.78
PSED ELG	477 (87.4)	440 (86.3)	0.06 (0.00, 0.12)	1.09 (-3.00, 5.17)	0.59 (-3.91, 5.08)	1.06 (0.70, 1.59)	0.03 (-0.22, 0.28)	0.80
Good Level of Development	393 (72.0)	355 (69.6)	0.07 (0.01, 0.13)	2.37 (-3.12, 7.86)	1.88 (-4.17, 7.92)	1.10 (0.81, 1.51)	0.06 (-0.13, 0.25)	0.54
	Mean (95% CI)	Mean (95% CI)	Correlation with baseline BAS3 ENC	Unadjusted mean difference (95% CI)	Adjusted mean difference (95% CI)	—	Hedges' g (95% CI)	P-value
ELG mean	1.90 (1.88, 1.91)	1.87 (1.84, 1.89)	0.11 (0.05, 0.17)	0.03 (0.00, 0.06)	0.03 (0.01, 0.05)	—	0.12 (0.04, 0.20)	0.02

Table 32: EYFSP outcomes for Cohort 2, sensitivity analyses omitting the mean baseline BAS3 ENC score per school from Cohort 1 as a covariate

Sensitivity analyses	Adjusted percentage point difference (95% CI)	Adjusted OR (95% CI)	Hedges' g (95% CI)	P-value
Maths ELG	6.38 (1.65,11.10)	1.56 (1.12, 2.18)	0.27 (0.07, 0.47)	0.01
Self-regulation ELG	0.72 (-3.48, 4.93)	1.08 (0.70, 1.66)	0.05 (-0.22, 0.31)	0.74
PSED ELG	0.70 (-3.81, 5.21)	1.07 (0.71, 1.61)	0.04 (-0.21, 0.29)	0.76
Good Level of Development	1.87 (-4.21, 7.95)	1.10 (0.80, 1.51)	0.06 (-0.14, 0.25)	0.55
	Adjusted mean difference (95% CI)	—	Hedges' g (95% CI)	
ELG mean	0.03 (0.00, 0.06)	—	0.12 (0.00, 0.24)	0.02

In sensitivity analyses, the analysis models for EYFSP Self-regulation and PSED ELG outcomes were repeated adjusting for baseline CSBQ scores for these domains rather than the BAS3 ENC. As expected, the correlation between the Self-regulation ELG and the self-regulation component of the CSBQ was higher (at 0.53, Table 33) than for the BAS3 ENC baseline score (0.03, Table 31). In the adjusted analysed, the magnitude of the Hedges' g effect size was increased (relative to when the BAS3 ENC was adjusted for) and favoured the control group (-0.34, 95% CI: 0.79 to 0.11, p = 0.14). Similarly, the PSED ELG was more highly correlated with each baseline CSBQ subscale score (absolute range 0.30 to 0.54, Table 33) than it was with the BAS3 ENC score (0.03, Table 31), and in the adjusted analyses, the Hedges' g effect size favoured the control group, though none of the differences were statistically significant.

Table 33: Sensitivity analyses for EYFSP outcomes for Cohort 2

Sensitivity EYFSP analyses	Correlation (95% CI)	Adjusted percentage point difference (95% CI)	Adjusted OR (95% CI)	Hedges' g (95% CI)	P-value
Self-regulation ELG, adjusting for CSBQ self-regulation	0.53 (0.48, 0.57)	-3.17 (-7.35, 1.02)	0.57 (0.27, 1.20)	-0.34 (-0.79, 0.11)	0.14
PSED ELG, adjusting for:					
CSBQ Sociability	0.33 (0.27, 0.38)	-1.98 (-6.56, 2.61)	0.81 (0.50, 1.32)	-0.13 (-0.42, 0.17)	0.40
CSBQ Externalising	-0.44 (-0.49, -0.39)	-2.19 (-6.38, 2.00)	0.78 (0.48, 1.26)	-0.15 (-0.45, 0.14)	0.31
CSBQ Internalising	-0.30 (-0.36, -0.24)	-1.55 (-6.52, 3.43)	0.85 (0.51, 1.43)	-0.10 (-0.41, 0.22)	0.54
CSBQ Prosocial	0.51 (0.46, 0.55)	-3.32 (-7.79, 1.15)	0.62 (0.33, 1.18)	-0.29 (-0.67, 0.10)	0.15
CSBQ Behavioural	0.54 (0.49, 0.58)	-3.05 (-7.53, 1.43)	0.63 (0.32, 1.24)	-0.28 (-0.69, 0.13)	0.18
CSBQ Cognitive	0.42 (0.36, 0.47)	-2.52 (-7.78, 2.74)	0.73 (0.37, 1.42)	-0.19 (-0.60, 0.21)	0.35
CSBQ Emotional	0.42 (0.36, 0.47)	-3.27 (-7.63, 1.08)	0.69 (0.42, 1.13)	-0.23 (-0.53, 0.07)	0.14

Teacher confidence

A total of 151 teachers (intervention, *n*=78; control, *n*=73) from 93 schools (intervention, *n*=47; control, *n*=46) completed the baseline survey and provided valid data for the teacher confidence questionnaire. Respondents reported that they had been teaching for an average of 11.8 years (SD 8.8, median 10) in the intervention group and 9.1 years (SD 7.4, median 7) in the control group. The average confidence summary score was 43.6 (SD 4.5) in the intervention group and 42.1 (SD 4.9) in the control group (Hedges' g effect size 0.31, 95% CI: -0.01 to 0.63). Responses are summarised in Table 34 and Table 35.

A total of 82 teachers (intervention, *n*=36; control, *n*=46) from 58 schools (intervention, *n*=28; control, *n*=30) completed the midpoint survey and provided valid data for the teacher confidence questionnaire, of which 30 had also provided baseline data (intervention, *n*=13; control, *n*=17).

A total of 105 teachers (intervention, *n*=56; control, *n*=49) from 65 schools (intervention, *n*=34; control, *n*=31) completed the endpoint survey and provided valid data for the teacher confidence questionnaire, of which 68 had also provided valid data at either baseline or midpoint (intervention, *n*=37; control, *n*=31); 48 at baseline only, three at midpoint only, and 17 at both.

Therefore, while it was the intention that, as far as possible, the same teachers would provide data at each time point, owing to the large movement of teachers only a relatively small proportion provided data at baseline and one follow-up time point (*n*=78).

Teachers in the intervention group reported greater confidence at both follow-up time points (in June 2022 to July 2022 after five to six months of intervention, and in June 2023 to July 2023 after the full 17 months of intervention). The unadjusted Hedges' g for the difference between the groups at midpoint is 0.58 (95% CI: 0.14 to 1.02) and at endpoint is 0.78 (95% CI: 0.38 to 1.17). The mean differences in confidence summary score, adjusted for baseline and years of teaching experience, are 2.1 (95% CI: -0.8 to 5.1, p = 0.16), and 3.1, (95% CI: 1.1 to 5.2, p = 0.003), respectively. The estimated adjusted Hedges' g effect sizes are 0.44 (95% CI: -0.17 to 1.04) and 0.69 (95% CI: 0.24 to 1.14).

Table 34: Your Confidence in Helping Children in Reception/Year 1 Learn Maths, part 1

I am confident in my knowledge of..., *n* (%)	Baseline		Midpoint		Endpoint	
	Intervention (*n*=78)	Control (*n*=73)	Intervention (*n*=36)	Control (*n*=46)	Intervention (*n*=56)	Control (*n*=49)
...what the children in my class know about maths when they enter						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	3 (3.8)	11 (15.1)	0 (0.0)	4 (8.7)	0 (0.0)	6 (12.2)
Neither agree nor disagree	13 (16.7)	10 (13.7)	5 (13.9)	8 (17.4)	3 (5.4)	5 (10.2)
Agree	47 (60.3)	40 (54.8)	25 (69.4)	28 (60.9)	35 (62.5)	32 (65.3)
Strongly agree	15 (19.2)	12 (16.4)	6 (16.7)	6 (13.0)	18 (32.1)	6 (12.2)
...reasonable maths goals for children in my class						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	0 (0.0)	2 (2.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Neither agree nor disagree	1 (1.3)	2 (2.7)	0 (0.0)	0 (0.0)	0 (0.0)	2 (4.1)
Agree	61 (78.2)	54 (74.0)	21 (58.3)	34 (73.9)	25 (44.6)	32 (65.3)
Strongly agree	16 (20.5)	15 (20.5)	15 (41.7)	12 (26.1)	31 (55.4)	15 (30.6)
...the best practices and strategies for helping children in my class learn maths						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	3 (3.8)	1 (1.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Neither agree nor disagree	14 (17.9)	25 (34.2)	0 (0.0)	6 (13.0)	1 (1.8)	2 (4.1)
Agree	57 (73.1)	43 (58.9)	25 (69.4)	31 (67.4)	22 (39.3)	32 (65.3)
Strongly agree	4 (5.1)	4 (5.5)	11 (30.6)	9 (19.6)	33 (58.9)	15 (30.6)
...national maths standards for children in my class						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	6 (7.7)	4 (5.5)	0 (0.0)	4 (8.7)	1 (1.8)	1 (2.0)
Neither agree nor disagree	15 (19.2)	12 (16.4)	2 (5.6)	4 (8.7)	2 (3.6)	5 (10.2)
Agree	45 (57.7)	44 (60.3)	23 (63.9)	29 (63.0)	25 (44.6)	29 (59.2)
Strongly agree	12 (15.4)	13 (17.8)	11 (30.6)	9 (19.6)	28 (50.0)	14 (28.6)
...the best ways to assess children's maths knowledge and understanding throughout the year						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	5 (6.4)	10 (13.7)	1 (2.8)	5 (10.9)	0 (0.0)	1 (2.0)
Neither agree nor disagree	21 (26.9)	30 (41.1)	1 (2.8)	8 (17.4)	1 (1.8)	7 (14.3)
Agree	47 (60.3)	30 (41.1)	26 (72.2)	23 (50.0)	28 (50.0)	34 (69.4)
Strongly agree	5 (6.4)	3 (4.1)	8 (22.2)	10 (21.7)	27 (48.2)	7 (14.3)

Table 35: Your Confidence in Helping Children in Reception/Year 1 Learn Maths, part 2

I am confident in my ability to..., *n* (%)	Baseline		Midpoint		Endpoint	
	Intervention (*n*=78)	Control (*n*=73)	Intervention (*n*=36)	Control (*n*=46)	Intervention (*n*=56)	Control (*n*=49)
...observe what children in my class know about maths						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	1 (1.3)	0 (0.0)	0 (0.0)	1 (2.2)	1 (1.8)	0 (0.0)
Neither agree nor disagree	1 (1.3)	2 (2.7)	0 (0.0)	1 (2.2)	1 (1.8)	0 (0.0)
Agree	53 (67.9)	56 (76.7)	24 (66.7)	30 (65.2)	24 (42.9)	31 (63.3)
Strongly agree	23 (29.5)	15 (20.5)	12 (33.3)	14 (30.4)	30 (53.6)	18 (36.7)
...incorporate maths learning into other areas of the curricula and school life						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	5 (6.4)	4 (5.5)	0 (0.0)	2 (4.3)	0 (0.0)	1 (2.0)
Neither agree nor disagree	10 (12.8)	8 (11.0)	1 (2.8)	6 (13.0)	2 (3.6)	2 (4.1)
Agree	50 (64.1)	52 (71.2)	22 (61.1)	28 (60.9)	27 (48.2)	34 (69.4)
Strongly agree	13 (16.7)	9 (12.3)	13 (36.1)	10 (21.7)	27 (48.2)	12 (24.5)
...plan activities to help children in my class learn maths						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	0 (0.0)	2 (2.7)	0 (0.0)	1 (2.2)	0 (0.0)	0 (0.0)
Neither agree nor disagree	8 (10.3)	8 (11.0)	1 (2.8)	1 (2.2)	0 (0.0)	0 (0.0)
Agree	59 (75.6)	52 (71.2)	18 (50.0)	32 (69.6)	21 (37.5)	32 (65.3)
Strongly agree	11 (14.1)	11 (15.1)	17 (47.2)	12 (26.1)	35 (62.5)	17 (34.7)
...further children's maths knowledge when they make spontaneous maths comments/discoveries						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	1 (1.3)	5 (6.8)	1 (2.8)	0 (0.0)	0 (0.0)	0 (0.0)
Neither agree nor disagree	6 (7.7)	7 (9.6)	0 (0.0)	3 (6.5)	0 (0.0)	3 (6.1)
Agree	52 (66.7)	51 (69.9)	15 (41.7)	30 (65.2)	22 (39.3)	28 (57.1)
Strongly agree	19 (24.4)	10 (13.7)	20 (55.6)	13 (28.3)	34 (60.7)	18 (36.7)
...make sense of children's confusions when they learn maths						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	4 (5.1)	3 (4.1)	0 (0.0)	1 (2.2)	1 (1.8)	0 (0.0)
Neither agree nor disagree	10 (12.8)	19 (26.0)	1 (2.8)	4 (8.7)	1 (1.8)	4 (8.2)
Agree	51 (65.4)	48 (65.8)	23 (63.9)	35 (76.1)	28 (50.0)	29 (59.2)
Strongly agree	13 (16.7)	3 (4.1)	12 (33.3)	6 (13.0)	26 (46.4)	16 (32.7)
...translate assessments into curriculum plans (i.e. turning assessments of children into next steps for learning)						
Strongly disagree	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Disagree	3 (3.8)	4 (5.5)	0 (0.0)	1 (2.2)	0 (0.0)	0 (0.0)
Neither agree nor disagree	11 (14.1)	22 (30.1)	3 (8.3)	6 (13.0)	2 (3.6)	6 (12.2)
Agree	54 (69.2)	42 (57.5)	22 (61.1)	32 (69.6)	27 (48.2)	31 (63.3)
Strongly agree	10 (12.8)	5 (6.8)	11 (30.6)	7 (15.2)	27 (48.2)	12 (24.5)
Confidence summary score, mean (SD)[a]	43.6 (4.5)	42.1 (4.9)	47.3 (4.2)	44.5 (5.1)	49.3 (4.3)	(4.3)

[a]*Summary score ranging from 11 to 55, and a higher score indicates greater confidence.*

Implementation and process evaluation results

Summary of IPE data collected by method

Surveys

Survey data were collected at the following time points throughout the evaluation:

- School-level usual practice surveys were issued at three time points: i) October 2021 to December 2021; ii) October 2022; and iii) June 2023 to July 2023. These surveys collected data on school-level usual practice and were completed by the headteacher, a member of the SMT, or the TEEMUP lead contact. The same person at each school may or may not have completed each survey at each time point.

- Teacher-level usual practice surveys were issued to nominated Reception and Year 1 in all schools twice: i) October 2021 to December 2021; and ii) June 2023 to July 2023. These surveys collected data on teacher-level usual practice. The nominated teacher in each school may have changed between time points due to staff changes/turnover.

- Intervention delivery surveys were issued to nominated Reception and Year 1 teachers in intervention schools at two timepoints: i) June 2022 to July 2022 following six to seven months of intervention delivery; and ii) June 2023 to July 2023 after 16–17 months of intervention delivery. The nominated teacher in each school may have changed between time points due to staff changes/turnover.

Table 36 summarises all IPE-related surveys sent to participating schools over the course of the trial including the number of valid responses received for each by allocation. Reminder emails were sent to non-responders to encourage completion at each time point. Survey responses were cleaned at each time point to remove invalid responses; where duplicate responses were received the first complete response from the correct intended respondent was used for analysis. Only valid responses were analysed.

Table 36: Summary of valid responses to IPE surveys at each time point by allocation

Survey type / Intended respondent	Control/intervention (total)			
	October 2021 / December 21*	June 2022 / July 2022	October 22	June 2023 / July 2023
School-level usual practice / Headteacher, SMT, or TEEMUP lead contact	37/42 (79)	NA	28/30 (58)	27/25 (52)
Teacher-level usual practice survey / Nominated Reception/Year 1 teachers**	73/76 (149∞)	NA	NA	55^/57 (112)
Intervention delivery (inc. cost questions) / Nominated Reception/Year 1 teachers	NA	NA/39× (39)	NA	NA/57Δ (57)

**Administered prior to randomisation.*
*** Prior to randomisation, all schools were required to nominate the Reception and Year 1 teachers who would receive the intervention should the school be allocated at random to receive it and to be a point of contact for the classes taking part within the school.*
∞ 149 valid responses were received from 97 schools.
^ 55 valid responses were received from 34 control schools.
Δ 57 valid responses were received from 34 intervention schools.
× 39 valid responses were received from 27 intervention schools.
NA=not applicable.

Interviews

It was originally planned that up to five relevant staff members (Reception and Year 1 teachers, SMT staff members) from seven schools would be recruited to participate in interviews, four schools forming 'longitudinal' case studies (four interviews between Summer Term 2022 and Summer Term 2023) and three schools forming case studies to investigate intervention delivery in schools demonstrating different levels of compliance (two interviews in Autumn Term 2022 and

Summer Term 2023). Recruitment to the interviews was challenging given the ongoing issues of staff illness/retention/time due to the COVID-19 pandemic.

A total of 23 interviews were conducted with 15 staff across nine schools (three longitudinal and six compliance) between May 2022 and July 2023 (Table 37). Interview participants included the schools' nominated Reception and/or Year 1 teachers, SMT, and/or maths leads (it is not possible to provide the proportion of these staff who participated in interviews as counts are below 5 and would be disclosive).

Table 37: Summary of schools participating in interviews by compliance

School level interview ID	Longitudinal or by compliance	Total no. of interview participants at school	Total no. of interview time points completed	Intervention compliance	
				Cohort 1	Cohort 2
01	Longitudinal	1	1	Non-compliant	Good
02	Longitudinal	2	3	Non-compliant	Good
03	Longitudinal	2	2	Non-compliant	Good
04	By compliance	2	2	Good	Good
05	By compliance	1	2	Good	Good
06	By compliance	1	1	Good	Good
07	By compliance	1	2	Good	Good
08	By compliance	2	2	Good	Good
09	By compliance	2	1	Non-compliant	Good

The trial statistician purposively selected 20 intervention schools that had not withdrawn from intervention delivery by Spring Term 2022 (a description of sampling criteria is provided within the 'IPE design' section of this report). Five staff members across three schools provided consent and participated in the longitudinal interview component. All five staff were interviewed in Summer Term 2022, two staff participated in Autumn Term 2023, none in winter/Spring Term 2023 and three in Summer Term 2023. Reasons for partial participation in all longitudinal interviews included staff leaving the school and their replacement not wanting to participate, school withdrawal from the intervention, or staff being unresponsive to the evaluation team contact at the time.

The delivery team identified 14 schools that met the 'good' compliance criteria in Autumn Term 2022. Excluding the schools that had already been invited to, or were participating in, the longitudinal case studies, the evaluation team emailed staff within eight of these schools and invited them to participate in two interviews, the first in Autumn Term 2022 and the second in Summer Term 2023. Four staff within three schools provided relevant consent and participated in an interview in Autumn Term 2022 as requested. In Summer Term 2023, three of these staff members from two of these schools completed the requested follow-up interview.

Staff within 13 schools that did not meet the 'good' compliance criteria in Autumn Term 2022 were invited to participate in interviews. Six staff within three schools provided consent, with five participating in the initial interview in Autumn Term 2022 (one staff member did not participate in Autumn Term 2022 but did in Summer Term 2023). In Summer Term 2023, three teachers within two schools completed the follow-up interview. Reasons for partial participation in these interviews included staff leaving the school and replacement staff not wanting to participate, staff being too busy to participate at the time, or staff being unresponsive to contact from the evaluation team; all of these factors were likely exacerbated by the COVID-19 pandemic.

The total number of schools randomised to the intervention was 47, so the nine schools that participated in the interviews represents 19%. Table 38 provides a comparison of school-level characteristics between all schools randomised to the intervention group and the group of schools that participated in the IPE. IPE sample schools had a slightly higher proportion of EAL pupils and pupils eligible for FSM than the overall randomised sample.

Table 38: Comparison of characteristics between randomised intervention schools and IPE schools

School level (categorical)	Intervention (randomised, *n*=47)	Intervention (IPE sample, *n*=9)
	Count (%)	Count (%)
%FSM[a]		
≤16%	22 (46.8)	3 (33.3)
>16%	25 (53.2)	6 (66.7)
%EAL[a]		
≤8%	23 (48.9)	4 (44.4)
>8%	24 (51.1)	5 (55.6)
Rural urban classification		
Urban	25 (53.2)	5 (55.6)
Rural	22 (46.8)	4 (44.4)
Participation in NCETM	25 (53.2)	6 (66.7)
School level (continuous)	Mean (SD)	Mean (SD)
%FSM[a]	20.0 (10.2)	23.4 (10.9)
%EAL[a]	18.2 (22.4)	22.1 (21.1)

[a] *Based on data provided by schools at the start of the evaluation, dichotomised for the randomisation at the median level for the recruited schools. FSM based on % of pupils eligible for FSM at any time during the past six years.*

To document the experience of delivering the TEEMUP PD, the Oxford delivery team participated in two online interviews/focus groups; the first in Autumn Term 2021–2022 following the bulk of the PD session delivery, and the second in Autumn Term 2022–2023 following intervention delivery.

Observations

Observations of the face-to-face TEEMUP PD training workshop were conducted to allow the evaluation team to gain a better understanding of the intervention. The evaluation team observed the following TEEMUP PD training:

- Full-day Sessions 1 and 2 (Spring Term 2022).
- Half-day Session 9 (Spring Term 2022).

It was also intended to observe Session 10 (scheduled for Summer Term 2023), however, this was not possible due to staff illness in the evaluation team.

Usual practice

This section explores the IPE results in relation to usual practice for early maths and maths CPD across all participating schools prior to randomisation (baseline) and explores what usual practice was for all participating schools during the trial. This section was informed by data from interview and survey data. The IPE research questions that are addressed in this section are:

4. What is 'usual practice' in all schools and has this changed in schools that have received the intervention?
    4.1 What is teacher's usual maths practice and has this changed in schools after receiving the TEEMUP PD?

4.2 How do teachers usually engage with children's families and the home learning environment and has this changed as a result of the TEEMUP PD?
4.3 How, if at all, do teachers work collaboratively within and across schools, and has this changed as a result of the TEEMUP PD?
4.4 How frequently, if at all, do teachers use formative feedback within their usual practice, and has this changed in schools that received the TEEMUP PD?
4.5 How frequently, if at all, do teachers use specific tools to reflect on their practice, and has this changed in schools that received the TEEMUP PD?
4.6 What is usual practice in relation to the transition between Reception and Year 1, and has this changed in schools that received the TEEMUP PD?

6. Is each stage of the intervention logic model supported by evidence and learning from the IPE?

CPD

Teachers' usual practice was ascertained at the start of the trial through questions about their experiences of CPD. Reception and Year 1 teachers were asked, on average, how many hours CPD they would normally complete in an academic year. Teachers reported an average of 21.1 hours of CPD per year across the two groups (intervention, 22.3 hours; control, 19.7 hours).

From the intervention group, 49% (45/76) of responding Reception and Year 1 teachers said that they had taken part in maths CPD in the 2020–2021 academic year, in comparison with 47% (34/73) of responding teachers within the control group. There was a wide range of CPD mentioned by respondents; recurring responses included Karen Wilding, White Rose Maths, and Maths Mastery training, EYFS maths and in-house school training in the form of staff meetings or classroom observations.

A high proportion of Reception and Year 1 respondents from both groups reported completed non-maths CPD in 2020–2021; intervention 82% (62/76) and control 82% (60/73). Common barriers for accessing CPD included the cost or budget restraints, difficulties getting staff classroom cover, and pandemic restrictions.

In total, 72% of respondents to the school-level usual practice baseline survey reported that their school planned to complete some form of CPD (other than TEEMUP) for EYFS/Key Stage 1 in the 2021–2022 year (the first year of the trial); intervention 74% (31/42) and control 70% (26/37). Additionally, 58% (46/79) of schools were also planning on implementing programmes or initiatives, in any area of EYFS/Key Stage 1 during the first year of the trial (2021–2022); intervention 50% (21/42) and control 68% (25/37).

Within all school-level usual practice surveys, the respondent was asked about their school's participation in specific CPD programmes, which were also providing maths CPD to Key Stage 1 teachers between the trial period, 2021–2023. These were initiatives/programmes available to schools at the time, which aimed to improve maths' attainment.

- **NCETM Maths Hubs Programme.** Schools can join one of 40 Maths Hubs Programmes, which run face-to-face maths CPD and online support.
- **NCETM Mastering Number.** Teachers from Reception, Year 1, and Year 2 are given face-to-face and online training and resources to support children's fluency and flexibility with number.
- **White Rose Maths:** Offers a wide range of maths CPD and resources for teachers.

While recognising the differing value and quality of these CPD initiatives/programmes (discussed further within the 'Limitations' section below) for the purposes of analyses, a school was considered to be taking part in these programmes if they completed at least one of the surveys and responded 'Yes' to the relevant question at any time point. In total, 64/93 (68.8%) of randomised schools reported participating in at least one of these programmes during the trial period (intervention, *n*=33/47, 70.2%; control, *n*=31/46, 67.3%). (NB. Actual participation rates may have been higher but not reported by the school to the evaluation team during the trial period.)

Table 39 presents a breakdown of such programme participation throughout the trial period 2021–2023, by randomised group. Control schools reported higher participation in NCETM Maths Hubs Programme and White Rose Maths in comparison to the intervention group, whereas the intervention group reported higher participation rates in NCETM Mastering Number during the course of the trial.

Table 39: Summary reported participation in other maths CPD programmes, by randomised group during the trial period

Maths CPD programmes	Randomised group		Total (*n*=93)
	Intervention (*n*=47)	Control (*n*=46)	
NCETM Maths Hubs Programme training	24 (52.2)	27 (60.0)	51 (54.8)
White Rose Maths PD training	10 (21.2)	12 (26.0)	22 (23.6%)
NCETM Mastering Number	18 (38.2)	16 (34.7)	34 (36.5)

Teaching practice

Reception and Year 1 teachers in all schools were asked about their usual practice in Autumn Term 2021 and again at the end of the trial in Summer Term 2023. This included how frequently teachers used formative assessment and reflective practice, and how often they considered ways to improve pupils' transition between Reception and Year 1. These data are presented in Appendix F Table 1.

Practices were similar between allocated groups at the start of the trial with 65% of teachers using formative assessment weekly to monitor and inform teaching. The use of tools/scales to reflect on teaching strategies was used less frequently by all teachers, with a third of teachers reporting that they used them weekly or monthly, and a third of teachers 'never' using them. Around half of all responding teachers thought 'often' about how they could improve the transitions of pupils into their new year group.

At the end of the trial, a greater proportion of teachers in intervention schools used formative assessment to monitor and inform teaching weekly in comparison to teachers within the control group (intervention, 70.2%; control, 54.5%). Such observed differences between allocated groups at the end of the trial could be indicative of teachers in intervention schools continuing to implement the formative assessment ideas/examples provided as part of the TEEMUP PD to support teachers in measuring maths to monitor child progress and inform their planning. The control group reported a reduction in weekly use of formative assessment by 11% in comparison to baseline levels (Autumn Term 2021, 65.8%; Summer Term 2023, 54.5%).

Teachers in the intervention group also reported using tools/scales to reflect on their teaching more frequently at the end of the trial in comparison to the control group; the data illustrated fewer teachers in trial schools reporting 'never' using such tools/scales following delivery of the TEEMUP PD (intervention, 12.3%; control; 25.5%). As part of the TEEMUP PD, teachers were supported to make quality improvements by using self-reflection tools such as the Behaviour for Learning and the Improving Maths Practice scales. The observed differences in the survey data could be indicative of teachers in intervention schools using such tools. There was a small reduction in the proportion of teachers in the control group reporting they 'never' used such tools/scales between the start and end of the trial (Autumn Term 2021, 31.5%; Summer Term 2023, 25.5%); this may be due to control group teachers participating in other CPD over the course of the trial (as detailed above) or because of a variation in respondents between the two surveys; a limitation which is discussed in further detail below.

Teachers in intervention schools thought more frequently about how they could improve the transition of pupils into new year groups following the delivery of the TEEMUP PD. Teachers in intervention schools did so more frequently than the control group at the end of the trial too. The TEEMUP PD encouraged collaborative learning, particularly between Reception and Year 1 teachers, to support pupils' transitions into classrooms.

It is important to note when interpreting these results that not all survey respondents will have completed both surveys, namely, respondents who completed the survey at the start of the trial, may not have completed the end of trial survey and vice versa. This should be considered when making survey comparisons. There is the potential for response bias and this is acknowledged within the 'Limitations' section below.

At the end of the study, both control and intervention teachers were asked open questions about their usual maths teaching. Many teachers across both groups mentioned following the White Rose Maths scheme of work, with some others mentioning Power Maths and NCETM Mastering Number. In the intervention group, TEEMUP was mentioned by three respondents.

The themes that emerged from the survey responses across both groups regarding usual practice at the end of the trial were very similar. Most answers from across both groups indicated that maths was taught daily, although this ranged from one to five times a week, with lessons varying in length from 15 to 60 minutes. Most teachers across both groups described using a mixture of teaching approaches to deliver maths, including formal whole-class teaching, usually on the carpet, followed by paired, group, or independent work that focused on a practical activity, and some element of written work within the week. Manipulatives and practical resources seemed popular teaching tools across both groups, although the use of stories as maths prompts was much more popular in the intervention group and was the only noticeable difference in responses. Many schools also mentioned that, in addition to set lessons, maths was embedded into other areas of the school day; examples of this included registration, story time, and telling the time. Continuous provision was popular and similar activities were mentioned across both groups, for instance, using construction areas, home role play using counting of recipes, maths games, money, time, playdough, and sand and water tables. Several teachers across both groups said they directly planned in their continuous provision to match the formal maths teaching that week. However, continuous provision seemed more popular with Reception teachers, with several teachers mentioning they were no longer using it or were phasing it out by the end of Year 1. One intervention teacher mentioned using a maths zone as a direct consequence of the TEEMUP training, and another mentioned using board games taken from TEEMUP.

Staff from all nine schools participating in the IPE interviews expressed that they felt their usual maths teaching practice had changed in some way because of the TEEMUP PD. Staff interviewed from five out of nine schools noted that as a result of the TEEMUP PD they were now actively encouraging pupils to explore and interact with maths through discussion. Staff interviewed from four out of nine schools discussed how their usual maths teaching practice now frequently incorporated the use of games, books, and stories; a greater use of stories to teach maths among the intervention group was also evident in endpoint survey data. During interviews, one school discussed how their usual practice involved following the White Rose Maths scheme, however they had moved away from this approach once a week so that they could implement practice from the TEEMUP PD.

Staff from three out of nine schools reported that the TEEMUP PD had led to improvements in communication between the nominated Reception and Year 1 teachers. Staff spoke of a newly acquired '*shared language'* of maths practice, which fostered greater communication day-to-day, and aided discussion between teachers when transitioning pupils from Reception to Year 1. One school described how the Reception and Year 1 created a list of '*non-negotiable*' (School 02, Autumn Term 2022) maths skills that children needed once they transition from Reception to Year 1. Children's maths skills and abilities were noted by some interviewees as a prominent topic of discussion between nominated Reception and Year 1 teachers during year group transition meetings.

> *Maybe because we've been on the sessions together it's kind of maybe brought us closer so it's easier to communicate with each other, you know, we didn't necessarily work closer on the subject of mathematics before that. We did on many other things, just in general. But yes, I think it has helped those [transition] conversations, definitely and obviously look at the data that she has completed for her group and speak about their needs and the SEN [Special Educational Needs] children and so on. So it's definitely helped. We would have done some anyway but it's definitely helped. We definitely know the language that each other speaks a bit more now as a result of some of the things we discussed in TEEMUP. So I think it's helped.*

It's hard to quantify but certainly on a qualitative basis it's certainly played a part. (School 06, Autumn Term 2022)

Home learning environment

A common theme from the interview data was that teachers struggled to engage parents/carers and encourage learning within the home environment. Teachers felt it was more difficult to engage parents/carers with maths-related activities, as home learning was traditionally phonics-based, for example, reading and writing, and were therefore, more accepted and expected by parents/carers. Teachers also noted other barriers to engaging parents/carers with maths-related home learning activities, these included parents/carers own maths anxieties and misinformed views on what maths teaching, practice and learning involved, for example, '*parents still think that maths is a sheet of sums*' (School 04, Summer Term 2023), '*trying to change that whole, in their [parents/carers] minds, maths is learning to count straightaway…Get them away from that more formal getting a child to sit down and write numbers*' (School 02, Autumn Term 2022).

The TEEMUP PD aimed to facilitate teachers' effectively engaging with the children's home in their maths education, and provided materials designed to support teachers in doing so. Staff from seven out of nine schools interviewed expressed that they had increased efforts to include maths within the home learning environment because of the TEEMUP PD. Teacher-level usual practice survey data collected at the start (Autumn Term 2021) and end (Summer Term 2023) of the trial supported the interview data; overall, teachers in intervention schools engaged more frequently over the course of the school year with pupils' families to improve the home learning environment at the end of the trial in comparison to the control group. These data are presented in Appendix F Table 2.

Among control schools, there was a large observable decrease in the frequency of engagement with the home learning environment between the start and end of the trial (e.g. the proportion of teachers engaging every six months with the home learning environment increased from 17.8% at the start of the trial to 27.3% at the end; see Appendix F Table 2). Among intervention schools, the proportion of teachers reporting that they engaged weekly with the home learning environment increased slightly between the start and end of the trial (32.9% vs 36.8%), however, so did the proportion of teachers reporting that they engaged on a yearly basis (9.2% vs 14%).

Teachers were initially requested to provide information on how frequently they engaged with the home learning environment at the start of the academic year in Autumn Term 2021. The completion of this survey followed the 2020–2021 academic year, which saw many teachers and pupils move to remote learning due to school closures resulting from the COVID-19 pandemic. This context very likely impacted on how frequently teachers in all schools engaged with the home learning environment at the time. Lower baseline levels of engagement may have been reported by teachers in all schools had the survey been completed pre-COVID. Nevertheless, in comparison to the control group, survey data indicate that around two-thirds of responding interventions, schools sustained more frequent engagement with the home learning environment at the end of the trial, following the delivery of the TEEMUP PD.

Within interviews and surveys, teachers provided descriptions of the types of engagement they had with the home learning environment. Examples included sending accessible maths games home with improved/clear instruction and communication on how and why maths is being taught the way it is, creating home learning packs containing various maths activities (some of which were TEEMUP resources), linking maths tasks to phonics homework, ensuring that resources required to complete maths activities are sent home (such as dice), and inviting parents/carers into school for in-person maths workshops, maths cafes, or sessions. A small number of teachers in intervention schools (discussed by two out of nine schools who participated in interviews and specifically noted by four respondents to the Summer Term 2023 intervention delivery survey) considered that their participation in the TEEMUP PD coupled with increased efforts to engage with the home learning environment had in turn improved parental engagement. For example, when asked to reflect on their engagement with children's families and the home learning environment, and how, if at all, it changed as a result of TEEMUP PD, a survey respondent commented:

A significant change. One of my goals with my mentor was to hold workshops with parents. I held 3 workshops and across these I had 85% attendance from parents, including several SEND. The feedback

from parents has been incredibly positive and I will do this again. (Summer Term 2023, Intervention delivery survey respondent).

Summary of key findings and applicability to the logic model

- Prior to the TEEMUP trial, teachers across all participating schools reported completing an average of 21.1 hours of CPD per academic year.

- Of all participating schools, 68.8% reported undertaking other Key Stage 1 maths CPD during the trial period; participation in such CPD was similar between randomised groups. Within the original TEEMUP logic model (Figure 1), a component considered to be an enabling factor of the success of the TEEMUP PD was that control schools continue with usual practice and do not engage with any substantial maths PD for the duration of the study and control schools reported high levels. We acknowledge that the level of school/teacher engagement with other Key Stage 1 maths CPD was not collected as part of these surveys (as this was out of scope), schools supplied the name of the initiatives/programmes that they were involved with. Nevertheless, this does not compromise the design of the evaluation because both intervention and control schools reported similar amounts of other maths CPD.

- Reception and Year 1 teachers reported usual practice in relation to the use of formative assessment, reflection on their own teaching practice, collaborative working, and improvements to engagement with the home learning environment, and the transition children between year groups. Data were broadly similar at baseline across allocated groups. However, the use of formative assessment, teachers' reflection on their own practice thoughts regarding improvements to the transition of pupils into new year groups increased in frequency among the intervention group at the end of the trial.

- A dominant theme within interview data was that staff in intervention schools felt their usual maths teaching practice had changed in some way because of the TEEMUP PD.

- A dominant view in both interview and survey data was that staff in the intervention group reported an increase in the use of games, books, and stories to teach maths.

- Staff in intervention schools had increased efforts to incorporate maths activities within the home learning environment because of the TEEMUP PD. Intervention schools reported more frequently engaging with families to improve the home learning environment in comparison to at the start of the trial. Survey data saw an increase in the proportion of intervention schools reporting engaging with the home learning environment on a weekly basis between the start and end of the trial. This is despite reported baseline levels of engagement likely being higher as a result of changes to normal teaching arrangements during the academic year 2020–2021 due to the COVID-19 pandemic. An output and subsequent short-term outcome of the original TEEMUP logic model was that teachers engage further with families by sharing resources and activities and subsequently report that families are using the maths games/activities at home. These data support the logic model component that teachers implementing the TEEMUP PD engaged further with the home learning environment.

- A short-term outcome within the original logic model was that teachers work more collaboratively with staff in class and across Reception and Year 1. Interview data evidenced this within three schools; however, this was not a dominant view across all IPE participants interviewed. This may have been due to the COVID-19 pandemic, which meant the intervention was unable to support collaborative learning as planned due to the need to observe social distancing guidelines.

Fidelity and adherence

Data collected in relation to fidelity aims to describe the extent to which the TEEMUP PD was delivered as intended within intervention schools. This section also explores any issues with schools' adherence. The IPE research questions addressed within this section are:

1. Is fidelity to the TEEMUP PD being observed?
    1.1 Are participating teachers attending/accessing the available training?

1.2 Are the different components of the TEEMUP PD materials and resources (e.g. reflective self-assessment scales, planning framework, use of formative assessments) being used as expected?
1.3 Are participating schools engaging with the school's allocated TEEMUP PD mentor as expected?
1.4 What are the barriers and/or facilitators to teachers engaging with the TEEMUP PD training and the mentor?
1.5 What constitutes necessary conditions (enabling factors) for participating teachers and schools to engage with the intervention as intended?
1.6 Do outcomes vary in line with compliance?

3. What are different stakeholder viewpoints of the TEEMUP PD?
3.1 What are teachers' perceptions on the usefulness and quality of the intervention as a whole and its components e.g. training, mentor support/visits, maths practice, and Behaviour for Learning scales, resources/materials?
3.4 How can the TEEMUP PD programme be improved?

6. Is each stage of the intervention logic model supported by evidence and learning from the IPE?
6.1 On review, after experience of PD delivery, do the delivery team consider any changes to the logic model necessary?

7. What can be learned from the efficacy trial to inform an effectiveness trial?

Data to answer these questions was provided from intervention surveys, interviews with relevant school staff and the delivery team, and monitoring data.

TEEMUP PD training

Training attendance levels demonstrate that Reception teachers who taught Cohort 1 had particularly high attendance at the first nine training workshops delivered between January 2022 and May 2022 (attending a mean of 8.2/9) (Table 40). Year 1 teachers who taught Cohort 1 had slightly lower attendance (mean 7.2/9), but this was still above the seven out of nine sessions required to be considered a 'good complier' within the training attendance component of the CACE analyses. Reception teachers who taught Cohort 2 had lower attendance at the first nine training workshops (attending a mean of 6.9/9 sessions delivered between January 2022 and May 2022). Also see Table 17 and Table 25.

Table 40: Attendance of teachers within participating schools at TEEMUP PD training workshops as defined by compliance criteria

No. of TEEMUP PD sessions attended	Cohort 1 Mean (SD), minimum-maximum	Cohort 2 Mean (SD), minimum-maximum
Reception teacher	8.2 (1.6), 0–9	6.9 (3.3), 0–9
Year 1 teacher	7.2 (2.5), 0–9	N/A

N/A=not applicable.

The dominant view from all survey data and staff interviews was that the TEEMUP PD training workshops were very well received by teachers who considered them to be of high quality, interesting, and useful.

> *The TEEMUP training has been the most positive training I have received over the past few years. It is very relevant to everything I am doing in class and has also tied in really well with our whole-school development. After each workshop there were inspirational things to try in class. The reasoning and problem-solving and communication and collaboration were particularly useful.* (Summer Term 2022, Intervention delivery survey respondent)

Where teachers could not attend face-to-face workshops, they could attend a trainer-led online workshop or catch-up independently by watching a recorded workshop. On average, Reception teachers of Cohort 1 attended 83% of the sessions they attended at a venue, 11% online, and 6% via a recording; Year 1 teachers of Cohort 1 attended 72% of the sessions they attended at a venue, 22% online (some of these were trainer-led Zoom 'new teacher' sessions in Autumn Term 2022), and 6% via a recording; and Reception teachers of Cohort 2 attended 80% of the sessions they attended at a venue, 14% online (including the 'new teacher' sessions), and 6% via a recording.

Teachers highly rated the quality of the training workshops, with venue-based training workshops receiving the highest mean quality (mean 8.6/10 in both Summer Term 2022 and Summer Term 2023) when compared to those delivered online (7.1/10 in Summer Term 2022, 7.6/10 in Summer Term 2023). These data are presented in Appendix F Table 3.

Teachers who reported receiving training face-to-face and online considered the content of the training and its quality to be similar between the two formats. For training workshops delivered face-to-face in early 2021, the delivery team adapted their initial approach to school collaboration during the sessions in response to many school requests to keep staff in school-specific 'bubbles' due to the COVID-19 pandemic. Despite this, teachers in four out of nine schools interviewed stated that they preferred venue-based training workshops, which they considered to be more interactive and collaborative than those online. Teachers valued the opportunity to discuss implementing the resources within the classroom and share ideas and practice with staff from other schools, as well as feeling there was more opportunity to ask more questions in comparison to online.

> *I don't feel like there was a difference in the information [between venue or online training] but…it was the general ability to practice with other people I think. So face-to-face when we were in the room we were given the chance, we were given certain ideas that we could use in our classrooms and we were then able to practice that amongst ourselves, we were able to translate it*. (School 08, Spring Term 2023)

Teachers also expressed the view that venue-based sessions allowed them to better focus on the TEEMUP PD, and provided valuable time for planning and reflection, particularly when they were accompanied by other staff from their own school.

Facilitators and barriers to training attendance

Of the 47 schools randomly allocated to the intervention, nine (19%) withdrew from delivering the TEEMUP intervention (eight out of nine schools were retained in the evaluation and provided at least primary outcome data). Some schools cited multiple reasons for withdrawal. The most commonly reported reason was due to staffing issues or changes within the school, followed by the cost to the school of staff cover for teachers to attend the training, or general financial constraints, or teachers' other time commitments.

A dominant view within survey and interview data was that the frequency of the fortnightly training delivered in 2021–2022 was a barrier to attendance. At a school level, this barrier was due to needing to find regular cover for Reception and Year 1 teachers to attend, particularly during a time of high staff absences due to COVID-19, and at a teacher level, having to fit the training around their existing school commitments.

> *It is a big commitment. The idea and training provided was amazing at the beginning but ideally it was meant to be both the Reception teacher and myself going on the training and that is too much for a small school to cover. Then when the Reception teacher left, we were left short so nobody could attend. I think this aspect, the amount of time needed to complete this training, along with an already unmanageable workload, is something to be considered.* (Summer Term 2022, Intervention delivery survey respondent)

It was evident within Summer Term 2023 intervention delivery survey data that teachers who received the weekly catch-up training online in 2022–2023 also perceived the frequency of training to be too intense. This weekly catch-up training was ran by the delivery team in response to high levels of staff turnover observed throughout the trial.

> *Weekly seminars and weekly prep work/action points were difficult to maintain and keep up with in an already demanding role*. (Summer Term 2023, Intervention delivery survey respondent)

The delivery team commented on the impacts of staff changes, illness of teachers (particularly during the COVID-19 pandemic), and the ongoing effects of the pandemic/economic crisis on staff stress and time. The delivery team acknowledged these as issues that could impact teachers' ability to engage with the TEEMUP PD. Based on their experiences of delivering the TEEMUP PD within the context of this trial, the delivery team created a revised logic model. Additional external moderators were added to reflect these points (please see Appendix E Figure 1 for the delivery team's revised logic model).

Schools that did not cite teaching cover as a barrier to training attendance explained that they did not need to outsource cover, as this was provided by existing higher level teaching assistants, suggesting having such staff was a facilitator to attendance.

Although noted by fewer teachers, having to travel to a venue was also considered a barrier to attending the training. For some teachers, the venues were quite a distance from their home or school. In such cases, teachers found attending an online session to be more convenient.

> *I needed to do a combination of at venue and online as the intense nature of the workshops (every 2–4 weeks) and being an hour drive away from home meant that I could not do in-person fortnightly. The combination of the sessions worked well.* (Summer Term 2022, Intervention delivery survey respondent)

Suggested improvements to the TEEMUP PD training

Teachers made several suggestions as to how barriers to training could be removed. These suggestions included that training workshops could be more spread out throughout the entire school year, and there was appetite from some teachers for '*more workshops in the second year but with full funding*' (Summer Term 2023, Intervention delivery survey respondent). Teachers felt that doing so would improve attendance and allow more time for teachers to consider how to implement the information and resources (discussed further in the 'Implementation' section below).

> *Spread the workshops out so that there is the opportunity for greater reflection and implementation between each workshop…fortnightly was unrealistic for this given everything else going on in a school*. (Summer Term 2023, Intervention delivery survey respondent)

> *Very useful ideas and suggestions that I could put into practice almost straight away—this was useful. However, there was a lot of information with the workshops being every 2 weeks and at times there seemed very little time to catch your breath, evaluate and reflect the workshop before moving on*. (Summer Term 2022, Survey respondent)

Additionally, several teachers commented that '*some sessions were a little bit repetitive'* (Summer Term 2023, Intervention delivery survey respondent) despite being '*very professional, clear and focussed*' (Summer Term 2023, Intervention delivery survey respondent) and as such could be condensed to reduce the overall number. One interview participant stated:

> *It was a nice day [the final training workshops in 2022–2023] to kind of reflect and look at everything that we've done. I know there might have been a bit too much reflecting because for about 30 minutes we re-watched a video from one of the workshops we've done before because it was the exact same video…I think there were a few tired faces looking at that. I think it's just nice to get together and just think about a subject because you don't ever get time to do that like really focus on maths or about behaviour. (School 07, Summer Term 2023)*

The repetition of training workshop content was also noted among teachers who attended online sessions.

> *We ended up doing most of the sessions online rather than in person. We were asked to watch videos before the sessions. Both elements were useful but I felt a lot of the information in the video was then repeated in the online session*. (Summer Term 2023, Survey respondent)

The delivery team acknowledged that some schools struggled to provide cover for nominated teachers to attend the TEEMUP PD training workshops, this was in part due to COVID-19 and the strain that COVID-19-related staff absence placed on school budgets.

> *People would phone me up saying, 'oh we're only sending one teacher because we haven't got enough budget to cover because we've got other sickness' and it was partly COVID because they were using a lot*

over COVID sickness. But they just said, we can't afford to send two people because we've used our cover for this month. (Delivery team interview, Autumn Term 2022)

The delivery team recognised that the 2021–2022 TEEMUP PD delivery schedule was condensed by the trial's timescales (i.e. pupil baseline and post-test assessments) and recognised the intensity of the weekly online catch-up sessions in 2022–2023. The delivery team suggested that should the TEEMUP PD be scaled up for an effectiveness trial, delivery would start in the Autumn Term, as opposed to January, and be spread over the two academic years as '*it seems to be really important to actually have those face-to-face sessions in the second year*' (Delivery team interview, Autumn Term 2023). Should the intervention be scaled up to an effectiveness trial, the delivery team suggested they would offer two full days and eight half-days of fortnightly sessions plus a consolidation workshop at a later date, totalling 50 hours (as opposed to the current delivery model of two full days and seven half-days of fortnightly workshops with an additional session, totalling 50 hours). A change to the training structure was included in the delivery team's revised logic model (Appendix E Figure 1).

The delivery team also recognised that travel to venues was a barrier to engagement, as teachers were reluctant to travel long journeys and the SMT were '*concerned about the additional time and stress that gives their teachers*' (Delivery team, Autumn Term 2023). The delivery team suggested that the future structure of hubs could be changed so that any future trial would have smaller clusters of schools, which would reduce teachers' travel times.

The delivery team acknowledged that it became clear to them during the TEEMUP PD sessions that allowing more time for planning was beneficial to the attending teachers.

TEEMUP mentor visits

As described in the TIDieR table (see Table 3), the intervention offered schools a minimum of three specialist needs-based coach/mentoring sessions, with more support if required. Here, the mentor would assist in supporting schools to implement changes following workshops, adapting the approaches to suit the school's context and children/families, and getting other staff (e.g. teaching assistants) involved. Mentor sessions predominantly took place in the 2022–2023 academic year, following the delivery of the TEEMUP PD training workshops in 2021–2022.

Table 41 details participating school's compliance to the mentor component of the compliance criteria. Compliance was higher among Reception teachers than Year 1 teachers, but overall, was good for most of the 38/47 randomised schools that did not withdraw from intervention delivery.

Table 41: School mentor compliance

Face-to-face mentor visits and preparedness	Cohort 1 *n* (%)		Cohort 2 *n* (%)
	Reception teacher	Year 1 teacher	Reception teacher
Good	31 (81.6)	27 (71.1)	30 (78.9)
Minimal	4 (10.5)	7 (18.4)	3 (7.9)
Non-compliant	3 (7.9)	4 (10.5)	5 (13.2)

Mentors were viewed extremely positively by the vast majority of teachers who responded to surveys or participated in interviews. Mentors were described by teachers as providing bespoke and tailored support in response to the needs identified by the school. Within survey and interview data, teachers commonly described their TEEMUP mentors as being supportive, helpful, understanding, very responsive to contact without being overpowering, with significant expertise, and many ideas on how to improve practice within the classroom.

Broadly across all responding schools, teachers described that TEEMUP mentors provided supportive expertise on how to improve and embed maths into the daily classroom routine by developing and implementing 'change plans', and applying knowledge acquired during TEEMUP PD into practice, signpost to TEEMUP resources, and/or make changes to the indoor and outdoor physical environment, for example, displays, placement, and availability of activities and resources. Some schools detailed that the needs-based support they received included advice on:

- improving engagement with parents/carers and the home learning environment;
- targeting practice to pupils who require additional support;
- improving behaviour for learning; and
- preparing and/or assisting in the delivery of training/information sessions to other staff within the school.

One interviewee described the contact and support they had with their TEEMUP mentor.

> *Every time [the mentor] came in she saw part of my lesson and a part of how I was really showing the TEEMUP kind of stuff. Then she would go and see part of the [R]eception maths going on and then we would have a good hour and a half chat, at least each, with her separately and a bit together kind of thing and she would give us more information or information to build upon what we've done. But we would also be able to talk about our change plans and what we'd specifically been focusing on. So that was really, really useful but equally I also emailed [the mentor] a couple of times in between sessions. Sometimes it was to let them know how something had gone. So once she was talking to me about how we could us the PD stuff and share it with TAs [teaching assistants]. So she spent a lot of time, so I would then email her and say this has worked really well or, you know, a little bit worried about this have you got anything more. Then other times it was just literally like questions: I'm struggling to think of something with this, have you got any ideas? So it was really supportive but not overpowering. I've done things before where you're constantly being almost bombarded with emails from your mentor to say what are you doing now? what are you doing now? How can I help? Can I do this? Whereas it was very much I knew she was there and I could email her at any point but I wasn't getting hassle of what are you doing? I haven't heard from you in two weeks, why haven't I heard from you? So that was really good. Sometimes I'd email her, you know, a couple of times in a week and then other times it might be for a little bit longer because I was using stuff that I didn't need her help with kind of thing.* (School 08, Summer Term 2023)

Survey respondents and interview participants described that they had email contact with their mentors, and that the mentors would visit their school for meetings and/or to observe their classroom. The Summer Term 2023 survey data indicated that TEEMUP mentors were viewed overwhelmingly positively by teachers, and this was also evident within interviews conducted within 2022–2023.

Supporting the data on mentors above, the quality and relevance/usefulness to practice of the mentor support was rated extremely highly by teachers throughout the duration of the intervention, with small increases seen between Summer Term 2022 and Summer Term 2023 intervention surveys (see Appendix F Table 4).

For 8/62 of the Summer Term 2023 survey respondents reported barriers that prevented them from engaging with their TEEMUP mentor over the course of intervention delivery were: being unable to be released from the classroom due to a lack of cover resulting from staff shortages; staff sickness; and/or budget limitations. A less commonly reported reason was 'other school priorities'. Similar barriers were discussed by interview participants.

During interviews, the delivery team made the following suggestions relating to mentor support should the intervention be scaled up for an effectiveness trial:

- Mentors could be recruited with different areas of expertise to support schools with different needs.
- Mentors could service one training hub containing up to ten schools.
- Holding mentoring sessions across the two years of intervention delivery, as opposed to only in the second year.

- Setting school expectations about mentoring from the start to increase school's engagement.
- An improvement to mentor training and quality assurance to maintain standards.

> *The training will have to be incredibly well managed to ensure that everybody has the ability to do what our three mentors have done this time.* (Autumn Term 23)

TEEMUP website

A minimum of eight school logins to the online knowledge base over the course of the whole intervention period was required for schools to be defined as a 'good' complier for within the CACE analysis, and a minimum of four logins was required to reach 'minimal' compliance. The minimum number of website logins for both Cohort 1 and Cohort 2 was ten, so this was above minimum compliance requirements (mean in Cohort 1 40.3; Cohort 2 39.3) (Table 42). These data demonstrate that schools engaged with the TEEMUP website.

Table 42: Number of school logins to the online knowledge base over the course of the intervention period

	Cohort 1 Mean (SD), minimum-maximum	Cohort 2 Mean (SD), minimum-maximum
Number of logins to website	40.3 (26.5), 10–120	39.3 (25.8), 10–120

Despite high engagement, a dominant view from survey and intervention data was that teachers found the website difficult to navigate, for example, that they were unable to find and/or access some resources, and that some video links were broken.

> *I really like the website resources. I just find the TEEMUP website hard to navigate and find things on.* (Summer Term 2022, Intervention delivery survey respondent)

Issues with the website were more frequently reported by teachers in the Summer Term 2022 intervention survey in comparison to the Summer Term 2023 intervention survey.[6] During interviews in Summer Term 2023, teachers discussed how the website had improved over the course of the TEEMUP PD.

> *Yeah I feel like looking from the beginning of the course, it wasn't as user friendly I would say. Whereas now it's a lot more accessible. You can find what you need and like I printed off the trajectories to sort of talk through my colleagues and I've been able to find them within seconds, whereas beforehand it was kind of like, where are they kept?* (School 09, Autumn Term 22)

Survey data also shows teacher's perceptions of the website improved between Summer Term 2022 and Summer Term 2023; the mean 'quality' rating increase from 8.0 to 8.5 and usefulness/relevance to teaching practice from 7.9–8.3 (see Appendix F Table 5).

Summary of key findings and applicability to the logic model

- The TEEMUP PD training and mentor sessions were very well received by teachers in the intervention group who considered them to be interesting, useful, and of high quality.
- Compliance data provided by the delivery team demonstrated that Cohort 1 teachers engaged with a mean of 75% of training workshops, and over 70% of teachers were considered to have had 'good engagement'

[6] The delivery team developed and refined the website throughout the first six months of intervention delivery in response to teacher and mentor feedback; the 'stable' version of the website was in place by September 2022.

with their TEEMUP mentor. These findings align to the training attendance component within the 'outputs' section of the original logic model.

- Teachers from four out of nine schools interviewed discussed how they preferred venue-based training workshops, although the online sessions enabled teachers to access the training when they could not attend it face-to-face. Offering online training and pre-recorded catch-up sessions enabled more teachers to engage with the TEEMUP PD, although they were not training formats that the delivery team had originally planned but were required due to the COVID-19 pandemic.

- The frequency of training (which was spread over five months) was a barrier to attendance, as schools struggled to arrange frequent cover, as was teachers' existing heavy workloads, particularly during COVID-19. Similar barriers were reported by teachers with regard to engaging with the TEEMUP mentor. Around 19% of intervention schools withdrew from delivering the TEEMUP PD.

- Both teachers and the delivery team considered that the training workshops could be spread out throughout the school year and even span two academic years.

- Schools accessed the TEEMUP website very frequently over the course of the intervention, which fulfils an 'enabling factor' component of the original logic model. Teachers reported issues navigating their way around the website, but their experience of using the website improved over the course of the intervention delivery.

Implementation

Data collected in relation to implementation aims to describe the extent to which the TEEMUP PD was implemented by schools as intended by the Oxford delivery team. The IPE research questions addressed within this section are:

1.2 Are the different components of the TEEMUP PD materials and resources (e.g. reflective self-assessment scales, planning framework, use of formative assessments) being used as expected?

2. To what extent is the TEEMUP PD implemented as planned within schools?
   - 2.1 To what extent do teachers implement the TEEMUP PD in their teaching practice?
   - 2.2 Have teachers adapted the intervention to make it more suitable for them, if so, how?
   - 2.3 What are the facilitators and/or barriers to teachers implementing the TEEMUP PD?
   - 2.4 What are the necessary conditions for teachers to implement TEEMUP PD into practice?
3. What are different stakeholder viewpoints of the TEEMUP PD?
   - 3.1 What are teachers' perceptions on the usefulness and quality of the intervention as a whole and its components e.g. training, mentor support/visits, maths practice, and Behaviour for Learning scales, resources/materials?
   - 3.4 How can the TEEMUP PD programme be improved?
6. Is each stage of the intervention logic model supported by evidence and learning from the IPE?
   - 6.1 On review, after experience of PD delivery, do the delivery team consider any changes to the logic model necessary?
7. What can be learned from the efficacy trial to inform an effectiveness trial?

Data to answer these questions was provided from surveys, interviews with teachers and the delivery team, and compliance data.

Use of the TEEMUP PD by teachers

Summer Term 2022 and Summer Term 2023 surveys requested nominated Reception and Year 1 teachers to rate the quality and the usefulness/relevance to practice of the TEEMUP PD components. Teachers were requested to indicate their response on a 10-point Likert scale (1 being 'poor quality'/'not useful', 10 being 'exceptional quality'/'extremely useful'). Full data are provided in Appendix F Table 6 with the mean score presented in Table 43 below. Teachers rated the following TEEMUP PD resources—Improving Maths Practice scale, Behaviour for Learning scale, learning trajectories, and formative

assessment materials—highly, receiving no less than a mean rating of 7.5. The most highly rated resource was the learning trajectories.

Table 43: Intervention survey data relating to the quality and relevance/useful of TEEMUP PD materials and resources over the TEEMUP PD, summary of mean responses

Likert rating, *n* (%)	Quality		Relevance/usefulness	
	Summer Term 2022 (*n*=39)	Summer Term 2023 (*n*=57)	Summer Term 2022 (*n*=39)	Summer Term 2023 (*n*=57)
Improving Maths Practice scale				
Mean (SD)	8.6 (1.1)	8.3 (1.5)	8.2 (1.4)	8.0 (1.7)
Behaviour for Learning scale				
Mean (SD)	8.4 (1.4)	8.0 (1.6)	7.8 (1.9)	7.7 (1.9)
Developmental progressions (learning trajectories)				
Mean (SD)	8.2 (1.5)	8.7 (1.4)	8.1 (1.5)	8.4 (1.7)
Formative assessment materials				
Mean (SD)	7.6 (1.5)	8.0 (1.5)	7.5 (1.5)	7.9 (1.5)

Table 44 presents data relating to participating school's 'mentor-determined evidence of change', which assessed each school's use of the TEEMUP PD tools and resources as part of the compliance criteria. These data indicate that most participating schools demonstrated 'good' or 'excellent' implementation of the TEEMUP tools or resources for both cohorts.

Table 44: Mentor-determined evidence of change as defined by compliance criteria

Mentor-determined evidence of change	Cohort 1 *n* (%)		Cohort 2 *n* (%)
	Reception teacher	Year 1 teacher	Reception teacher
Excellent	24 (63.2)	22 (57.9)	25 (65.8)
Good	8 (21.2)	6 (15.8)	5 (13.2)
Minimal	3 (7.9)	6 (15.8)	3 (7.9)
Non-compliant	3 (7.9)	4 (10.5)	5 (13.2)

The dominant view within interview and survey data was that the TEEMUP PD experience (e.g. training, scales, resources, and website content) was received positively by teachers.

> *I have found the programme to be a very positive experience. The face-to-face sessions were excellent and allowed sharing of ideas. The content of the PD was very interesting. The resources we were signposted to were great and the IMP [Improving Maths Practice scale] and BfL [Behaviour for Learning scale] have given me ideas for whole-school improvement in maths and across other areas of the curriculum.* (Summer Term 2022, Intervention delivery survey respondent)

There were no instances within interviews or surveys where teachers reported not having implemented anything from the TEEMUP PD. Ten survey respondents specifically highlighted that the TEEMUP PD provided a large volume of information which was '*a lot to digest*' (Summer Term 2022, Intervention delivery survey respondent), particularly alongside normal teaching commitments. As such, teachers implemented some of the TEEMUP PD components more frequently than others. The decision on what to implement was often in response to context, for example, adapting to the needs of the pupils/class, the time teachers had available, and suggestions from the TEEMUP mentor.

> *So obviously when you get training you don't take everything away from it and think, oh I'm going to use everything. One of the main things that I've taken from the training is the use of learning trajectories to show the development of certain topics over time. So I use them all the time now when planning different topics, you know, it's just part of my planning now. I do refer to the IMP [Improving Maths Practice scale] and sometimes the BfL [Behaviour for Learning scale] but not as much as the learning trajectories and then in terms of resources, I still do use a lot of the resources.* (School 05, Summer Term 2022)

As detailed in Table 43 above, the Improving Maths Practice scale, developmental progressions (learning trajectories) and the TEEMUP website resources received the highest mean ratings in quality and usefulness/relevance to practice from teachers within the intervention survey data. One interview commented on their prolonged use of the TEEMUP PD resources:

> *I've obviously looked through both the IMP [Improving Maths Practice scale] and the BfL [Behaviour for Learning scale] to begin with when I looked I thought, oh no I don't do that, I don't do that, you know, it was quite daunting to begin with. But as you use them more you can sort of see, yeah I do do that bit. They're really well broken down I think and kind of achievable sections.* (School 02, Summer Term 2022)

The developmental progressions (learning trajectories) were the most prominently discussed TEEMUP resource during interviews with teachers. Teachers within five out of nine schools interviewed described regularly using the developmental progressions (learning trajectories), particularly to facilitate planning.

> *We really do use…the trajectories in our planning. So just make sure the lessons are sequenced correctly and the correct level of support is there for the children that need it. We've actually adapted our long-term and medium-term planning so it follows the sequence of, you know, starting with subitising and then moving on from there. Just to get that early understanding of number.* (School 09, Autumn Term 2022)

Teachers reported 'slowing down' teaching in response to implementing the development progressions (learning trajectories) within their teaching. These resources encouraged teachers to '*address their [pupils] small steps*' (Midpoint survey respondent). Teachers described that the development progressions (learning trajectories) allowed pupils of all abilities to engage with the same activity by allowing them more time to investigate and explain their thinking.

Implementation of the Behaviour for Learning scale received some mixed feedback. One teacher described favouring the learning trajectories over the Behaviour for Learning scale, as they found the Behaviour for Learning scale challenging to implement among their high-needs class, whereas another school reported using the Behaviour for Learning scale frequently.

> *In terms of the BfL [Behaviour for Learning scale], I find that trickier to try and implement mainly because I think out of our like 39 [pupils], we've got 10 high needs and these are extreme needs and I find it really affects the other children because there's no general support staff.* (School 07, Autumn Term 2022)

> *I've done loads on self-regulation. We've made that really fun as well. Yeah we've made it into a Kylie moment. So we put our 'put your hand on your heart and tell me' and the children love it. We do that and they do actually say, I'm a ten and I say that's a really big number why are you feeling so sad? And it really helps them to come to terms with it, deal with it.* (School 08, Summer Term 2023)

Barriers and facilitators to implementation

Having support within the school was identified as the main necessary condition for teachers to implement components of the TEEMUP PD within their classrooms. Teachers highlighted that attending the TEEMUP PD with another teacher(s) from their school not only improved their experience of the intervention but meant they were able to implement the TEEMUP PD more effectively into practice. Teachers discussed the importance of having support from the headteacher and/or the school maths lead, who granted nominated teachers the freedom to implement the TEEMUP PD as necessary and purchase new resources.

> *So [name] he's SLT [senior leadership team] presently in charge of core and he's the one that I would go to immediately after the sessions and liaise with him about it and he's been really, really supportive and making sure that we're always doing everything, you know, checking in with us, do we need to do anything for TEEMUP, what's going on? He has been really good to us.* (School 05, Summer Term 2022)

One school discussed how the support of the maths lead increased over the duration of the TEEMUP PD as the improvements to children's school data were observed:

> *And I do think our maths lead generally is beginning to understand a lot more…if we [early years] don't get it right, it affects the whole…it's really important that we get it right and we try different things and therefore it will eventually filter through to his class. Yeah we have been supported in that sense really, haven't we?* (School 04, Summer Term 22)

> *He [maths lead] was very much for, 'yes carry on as you are', you know, he obviously sees all the data. He's very much data driven than perhaps we are and, you know, the data has put big smiles on his face. He's quite happy.* (School 04, Summer Term 2023)

It was apparent during interviews that in schools where the maths lead attended the TEEMUP PD training workshops it was easier for them to make changes/decisions regarding the implementation of the TEEMUP PD.

> *I think it's benefited me because I'm the maths lead for the whole school, so I can just come back [from TEEMUP PD training workshops] and I have the authority to change things if I want to. So I can go and be like, right this is now what we're going to be doing and I can come back and change it and I think that's had a huge impact. I think some schools don't always get, you know, they're not always that flexible in terms of they follow schemes and things like that and are quite rigid. Whereas I've been like, right [name] you're fine to do that next week, go on and have a go at doing it and put that in your planning. Okay we're just going to buy all these resources because we need them and we're going to use those. So I think that has had a huge benefit. I think, you know, if I hadn't been the maths lead and trying then to persuade someone else, the training might have been tricky. I'd say that was the main factor and just maybe the fact that I was new to [R]eception when we started it and [name] was new to [Y]ear 1, so we have never done it before and so learning things, we were like, why not, let's try it because we've never done it before. So I think new to the year groups and also being maths lead are maybe the biggest factors.* (School 05, Autumn Term 2022)

Having support from the wider school enabled teachers to adapt the intervention for use with older year cohorts. Teachers described running maths focused parent/carer workshops and/or delivering maths focused teacher training workshops for older year groups using TEEMUP resources, for example, Improving Maths Practice and Behaviour for Learning scales.

> *The resources that we've been signposted to are, I believe, high quality and they can be extrapolated to, you know, simple number recognition in [R]eception and in [Y]ear 1 obviously multiplication tables and beyond into [Y]ear 6.* (School 01, Autumn Term 2022)

The headteacher's support enabled teachers to implement the TEEMUP PD in schools where other maths schemes were been followed.

> *I actually have a scheme which obviously the school is following but after the problem-solving session with TEEMUP, I had a chat with the head and asked if I could move away from the White Rose [Maths] once a week so that we have a lesson that is purely reasoning and problem-solving and she was very happy for me to do that.* (School 03, Summer Term 2022)

However, having to implement other maths schemes was also recognised as a barrier to implementing the TEEMUP PD.

> *Quite often, Year 1 teachers have to teach a particular maths scheme (we teach White Rose [Maths]) and it can be difficult for Year 1 teachers to implement some of the TEEMUP content alongside teaching from a scheme.* (Summer Term 2023, Intervention delivery survey respondent)

The delivery team acknowledged the importance to the implementation of the TEEMUP PD from having the *'senior leadership to support and [recognise] the fact that they [teachers] are making changes'* (Delivery team interview, Autumn Term 2022). The delivery team felt that having schools' SLTs on board with the TEEMUP PD would increase engagement and help facilitate mentor visits, etc. They suggested that a potential adaptation to the TEEMUP PD training could be to include

these members of staff in all sessions. The delivery team's revised logic model incorporates this suggested change with the inclusion of a meeting specifically for the SLT in order to explain TEEMUP and the commitment required from schools (Appendix E Figure 1).

Being part of a MAT and having to follow a Trust's teaching expectations was highlighted as a barrier to implementation. Teachers described needing to adapt the TEEMUP PD to be able to implement it in a way which would be acceptable to the Trust:

> *From my point of view it [TEEMUP PD] doesn't align with what the Trust want. We've had specific training from educational executives within our Trust that have said things that I've directly thought go against some of the TEEMUP things which does make it challenging because then it's looking at, you know, I find it hard because I look at what TEEMUP is saying and think, okay how can we translate that in a way that is still acceptable? So it's following what I believe and what I've taken from TEEMUP but within the parameters that has been dictated by the Trust.* (School 08, Spring Term 2023)

Year 1 teachers were more likely than Reception teachers to describe being unable to implement the TEEMUP PD during the academic year 2021–2022. This was because the timing of the TEEMUP PD did not align with what was been taught in school at the time. However, these teachers did note that they were looking to implement the TEEMUP PD more fully in the next academic year.

> *I have found all the information extremely informative and have been very excited to start, however I have found that we are often learning about things at the wrong time of year…I have personally found it tricky trying to use the materials when they have been given to us at different times of the year. However, I am looking forward to starting the next academic year with all of the tools I have been given and really seeing the impact in July 2023 for my new [R]eception class!* (Summer Term 2022, Intervention delivery survey respondent)

A frequently reported barrier to implementation was a lack of time due to an existing heavy workload, which had increased due to the COVID-19 pandemic. Teachers valued the time out from the school day to attend the TEEMUP PD training workshops, which allowed them to focus on the content; however, they felt they would have benefited from having time out of the classroom aside from the TEEMUP PD training workshops. This time would have enabled them to reflect and consider how '*to absorb and integrate their learning from TEEMUP into their practice. It could feel a little overwhelming at times*' (Summer Term 2023, Intervention delivery survey respondent), given the volume of information shared. When asked how the TEEMUP PD could be improved, a frequently reported suggestion from teachers related to additional 'fully funded' time.

> *The TEEMUP PD has been very good this year but a bit overwhelming alongside teaching and leadership responsibilities. Lots of really useful information, guidance and resources but I feel I need time to process this and find ways to implement.* (Summer Term 2022, Intervention delivery survey respondent)

During the delivery team interview, the pressures on teacher's time was acknowledged. The delivery team explained that in response to these pressures, they had produced additional resources that teachers could use in the classroom that required less planning and adaptation.

Summary of key findings and applicability to the logic model

- The TEEMUP PD experience (e.g. training, mentor, scales, resources, and website content) was received positively by teachers. The TEEMUP PD provided an abundance of information and resources, which teachers considered to be of high quality and relevant to their practice. Attendance at training and engagement with mentoring sessions, website content, etc. were key 'outputs' from the logic model with the aim of facilitating positive 'short-term outcomes' such as increased teacher knowledge and confidence in teaching maths and increased pupil engagement with maths.

- Teachers frequently reported struggling to navigate the website particularly in the first year of the trial (2021–2022), although subsequent improvements were made to the website and noted by teachers. Teachers engaging with the website is considered an 'enabling factor' within the original logic model. Teachers being

unable to use the website effectively may have limited implementation of its contents, particularly during 2021–2022.

- Compliance data demonstrate that most participating schools had 'good' or 'excellent' implementation of the TEEMUP tools or resources, as determined by the TEEMUP mentor. Within IPE data, there was evidence that teachers implemented components of the TEEMUP PD within their practice to facilitate planning and to be responsive to children's needs, with reference to the developmental progressions (learning trajectories). These findings support an 'output' component of the original logic regarding teachers evidencing changes within the classroom using TEEMUP PD tools and resources.

- The decision on which components of the TEEMUP PD to implement was often in response to context, for example, adapting to the needs of the pupils/class, the time teachers had available, and suggestions from the TEEMUP mentor. This finding aligns itself with TEEMUP's need-based approach to support responsive teaching, and subsequently the rational for the TEEMUP intervention within the revised logic model (Appendix E Figure 1).

- Some schools adapted the TEEMUP PD resources for use with older cohorts, which was not part of the trial but has the potential to widen the scope and reach of the TEEMUP PD intervention and benefit more children. This was not specifically considered in the logic model as the trial focused on two year groups (Reception and Year 1) only.

- A necessary condition for implementation of the TEEMUP PD within schools was having the support of other school staff who were open to a change of practice. The headteacher and the school's maths lead were noted to be of particular importance. This finding supports an 'enabling factor' within the original logic model that highlights the importance of school management and leadership being supportive of study, collaboration, and changes in school.

- A frequently reported barrier to implementation was that teachers lacked time to consider how best to integrate the TEEMUP PD into their teaching. Other barriers to implementation included schools being part of a MAT and needing to follow a more prescriptive teaching framework and following other maths programmes. These are identified barriers to an 'output' from the logic model that teachers should be able to implement and evidence changes in classroom practice based on participation in TEEMUP PD activities.

- Some Year 1 teachers reported being unable to implement the TEEMUP PD during 2021–2022 as effectively as they would have liked, as the timing of the TEEMUP PD did not align with what was been taught in school at the time.

Perceived impact

Data collected in relation to stakeholder perspectives aims to describe how staff perceived the effectiveness of the TEEMUP PD in relation to the impact it had on their own ability and confidence to teach maths and self-regulation strategies, and also any observed impacts on the children as a result of any changes in practice. The IPE research questions addressed within this section are:

3. What are different stakeholder viewpoints of the TEEMUP PD?
   - 3.2 What is the perceived impact of the TEEMUP PD on teacher's maths practice, teacher's confidence in teaching children maths, and teacher's confidence in their own maths abilities?
   - 3.3 What is the perceived impact of the TEEMUP PD on children's maths outcomes and self-regulation? Are there any perceived differential intervention benefits among disadvantaged children?
   - 3.5 What is the perceived impact of the TEEMUP PD on teachers' practice and confidence in relation to children's self-regulation?
6. Is each stage of the intervention logic model supported by evidence and learning from the IPE?

The data sources were the interviews with school staff and intervention survey data.

Perceived impact on teachers

A dominant view among teachers was that they felt that participating in the TEEMUP PD had led to improvements in subject knowledge surrounding maths teaching practice. Teachers described feeling '*upskilled*' (Summer Term 2022, Intervention survey respondent), which they considered was now reflected in their teaching practice.

> *I found all 9 workshops very informative and I believe they have improved my subject knowledge hugely. The trainers own insights from her own classroom practice have also been relevant and helpful. I enjoyed learning about the research behind the PD and found that the sessions have provided lots of practical ideas to take back and try out in the classroom. Above all, I think the sessions have inspired me to want to provide better Maths provision for my children and have encouraged me to think about how to plan for lessons full of awe, wonder and discovery.* (Summer Term 2022, Intervention survey respondent)

Although improvements to maths teaching practice was dominant within the data, some teachers also reported improvements to their subject knowledge with regard to self-regulation.

> *The mentor support, workshops and resources/resource ideas provided has been extremely useful and I now feel much more confident and comfortable teaching Maths. I feel like since I have been on the course I have my love for maths has increased and this has rubbed off on the students. The course also supported me with other areas of teaching such as behaviour management.* (Summer Term 2022, Intervention survey respondent)

Within interviews, a minority of teachers specifically reported that their '*general knowledge around maths*' (School 07, Autumn Term 2022) had improved because of the TEEMUP PD.

Staff within many schools perceived some improvements to confidence in their maths teaching practice because of participating in the TEEMUP PD. Some teachers explained that they perceived improvements to their own personal confidence in teaching children maths.

> *A hundred percent it's changed my confidence in teaching maths…I do feel stronger teaching maths in EYS [early years stage].* (School 05, Summer Term 2023)

Other teachers reported that they perceived the TEEMUP PD had given them confidence to move away from current maths teaching approaches/schemes and to try new methods.

> *Confidence building, supportive, informative, in-depth. The PD has encouraged me to think about the teaching of maths rather differently. It has given me the confidence to challenge current procedures and try different methods/activity types. I appreciate the background information on research which has informed pedagogy. I liked that I was able to link back to research shown when introduced to activities/methods.* (Summer Term 2023, Intervention survey respondent)

> *Has my confidence changed? … not in the actual ability to teach it. My confidence has changed in the ability to break away from like following a scheme, you know, that kind of like not doing this, and this, and this but being able to slow it down and pace it to the needs of the children. Not like how we do and what we do it but the confidence has changed to say actually, no, I'm not going to do that. I'm going to wait and do something else. Like not saying you have to follow this and then this. So kind of use it rather than to follow religiously the pattern.* (School 04, Summer Term 2023)

Several teachers reported that the TEEMUP PD had given the confidence to 'slow down' their usual teaching approach.

> *I think it's given the confidence to slow down doesn't it, slow down. You don't feel you've got to rush through. Spend the time talking to them. Spend the time getting them to talk and knowing the benefits of that and I just think it's confident to slow down the pace isn't.* (School 03, Summer Term 2022)

Not all teachers reported improvements in their confidence, for example, one teacher discussed how they had observed positive improvements to their team because of the TEEMUP PD rather than within themselves.

> *I don't think it's necessarily had a massive impact on my confidence in teaching maths and my own ability in teaching maths but I do believe that it's had a positive impact on the [R]eception team and I feel confident moving forward that thanks to the resources you guys have signposted me to, I feel confident that it will have a positive impact on my team because, you know, your resources are very well sourced and thought through which is why I'm going to nicely magpie them or steal them. It's up to you whichever term you want to use!* (School 01, Summer Term 2022)

Perceived impact on pupils

Most survey respondents and interview participants described that they perceived that the TEEMUP PD had had a positive impact on both children's maths and self-regulation.

A dominant view among teachers following the first year of the TEEMUP PD was that pupils had enjoyed maths within their classes and had become '*more confident mathematicians*' (Summer Term 2022, Intervention survey respondent).

Teachers described a variety of improvements to pupils' maths development and learning including: reasoning; resilience when problem-solving; increase in the use of mathematical language when communicating mathematical understanding and in general; and a willingness to engage with maths as '*the children know that it is safe to verbalise their ideas and to discuss mathematical possibilities*' (Summer Term 2023, Intervention survey respondent) as '*any answer is accepted*' (Summer Term 2023, Intervention survey respondent).

Although evidenced within midpoint survey data, it was more common for responding teachers to report that pupil's '*maths data [had] improved from previous years*' (Endpoint survey respondent) within the endpoint survey (Summer Term 2023). Some, but not all teachers, attributed these improvements to directly to their implementation of the TEEMUP PD:

> *Maths scores using PUMA [Progress in Understanding Maths Assessment] were high this year. Children have had TEEMUP since [R]eception. Able to justify teaching maths how I want to!* (Summer Term 2023, Intervention survey respondent)

Other teachers acknowledged that TEEMUP PD ran in unison with changes to ELG Expectations statutory framework, which may have also contributed to pupils' improved maths outcomes.

> *Our maths expected levels at the end of the year have been much higher than in previous years; however, it is difficult to tell whether this is as a result of the change in Early Learning Goal expectations in the new statutory framework, or whether as a result of implemented changes following the CPD (or both).* (Summer Term 2022, Intervention survey respondent)

The impact of the TEEMUP PD on children's self-regulation was also a prominent theme within the endpoint survey data. For some teachers '*the most notable area [of observed impact] has been in self-regulation, this is because for this cohort this was the greatest need*' (Summer Term 2023, Intervention survey respondent). Some teachers reported positive changes to pupils' behaviour in the classroom through implementing the TEEMUP PD strategies.

> *Self-regulation has been a huge factor this year with many children in my class unable to self-regulate at all at the beginning of the year. Now, at the end of the year, all of the children can self-regulate to some degree and some just need a little adult support to do this. There has been a significant improvement in this.* (Summer Term 2023, Intervention survey respondent)

Teachers had mixed views as to whether the TEEMUP PD benefited some groups of pupils more than others. Some teachers reported that they felt the TEEMUP PD enabled them to better support pupils of all abilities to access and engage with the same maths activity.

> *I am able to support pupils of all abilities much more, grounded in evidence-based research*. (Summer Term 2022, Intervention survey respondent)

Whereas other teachers perceived that lower ability pupils benefited the most; particularly as the development progressions (learning trajectories) encouraged teachers to slow down their teaching and identify gaps in individual pupils' knowledge, enabling them to '*better support lower ability pupils*' (Summer Term 2023, Intervention survey respondent).

> *So my very low ability children last year I think they benefited because we really stripped back and we really back tracked along the trajectories and ironed out exactly where the gaps were. So I think the sort of children that would probably get swept along with the rest of them but really missed exactly what they do know or what exactly they don't know I think were more uncovered last year and then those gaps were filled that might not necessarily have been filled before they had a bit more understanding. So, yeah, hopefully, definitely the lower ability children I think it really did help.* (School 02, Autumn Term 2022)

One teacher described that they perceived the TEEMUP PD approaches to '*lift up the "lower ability" but do not stretch the "higher ability"*' (Summer Term 2022, Intervention survey respondent). Another teacher reported '*a really big improvement with the EAL children*' (Summer Term 2022, Intervention survey respondent) because of implementing the TEEMUP resources.

Summary of key findings and applicability of findings to the logic model

- Teachers generally considered that TEEMUP PD had upskilled their maths teaching practice and they perceived their confidence had improved, which were key 'short-term outcomes/mediators' listed in the logic model.
- Teachers perceived an improvement in their pupils' confidence in maths, which was a key 'short-term outcome/mediator' listed in the logic model with the aim of facilitating the longer-term outcome of improving maths attainment.
- Teachers also reported a positive impact on children's self-regulation/behaviour within the classroom.

Adherence to trial procedures

This section explores how the effectiveness trial's recruitment and pupil outcome data collection strategies, and attrition, may have affected the estimated impact of the TEEMUP PD. The IPE research questions addressed in this section are:

5. To what extent does the TEEMUP impact evaluation adhere to the proposed plan?
   - 5.1 Does the child recruitment and assessment process adhere to the plans proposed in the protocol?
   - 5.2 Any there any sample attrition effects? If so, how might that affect the estimates of the impact of the TEEMUP PD?

To address these research questions, we summarise data gathered within the impact evaluation.

Child recruitment

Detailed child recruitment figures are provided in the participant flow section of the impact evaluation results. The child recruitment process mainly adhered to the proposed plan, though not all randomised schools recruited a Cohort 2. Of the 93 randomised schools that recruited a Cohort 1, 73/93 (78.5%) went on to recruit a Cohort 2 in Autumn Term 2022. Of the remaining 20 schools that did not recruit a Cohort 2, two had fully withdrawn from the evaluation and 18 declined to recruit children into Cohort 2 but agreed to continue in the evaluation for Cohort 1. As the primary outcome data were collected for Cohort 1, not Cohort 2, the priority was to retain schools for Cohort 1 post-test.

As part of recruitment for Cohort 1 and Cohort 2, schools were asked to provide the number of parent/carer withdrawal forms returned (see 'Methods' section for details on the pupil recruitment strategy).

For Cohort 1 recruitment, 70/93 (75.3%) randomised schools provided the number of withdrawals forms returned (*n*=68; mean per school 1.0).

For Cohort 2 recruitment, 51/73 (69.7%) schools that recruited a Cohort 2 provided the number of withdrawals forms returned (*n*=30; mean per school 0.6).

Assessment administration

The baseline (prior to randomisation) and post-test BAS3 ENC assessments were administered by blinded, independent research assistants within each participating school. In total, independent research assistants visited:

- 93 schools at baseline (Cohort 1 only) and administered the assessment to 1,566 participating pupils (intervention, *n*=780; control, *n*=786).
- 91/93 randomised schools at post-test and administered the assessment for Cohort 1, of which 1,330 participating pupils' data were included within the primary analysis (intervention, *n*=664/780, 85.1%; control, *n*=666/786, 84.7%).
- 73/93 randomised schools at post-test and administered the assessment for Cohort 2, of which 1,001 participating pupils' data were included within the analysis (intervention, *n*=525; control, *n*=476).

Prior to their first school visits, research assistants were extensively trained by the evaluation team and an external educational psychologist with specialist knowledge of administering the BAS3 ENC to school-aged children. Either a senior member of the evaluation team or the educational psychologist observed each research assistant on one of their initial school visits to quality assure the administration of the assessment to check whether the administration closely followed the training and protocols (at baseline, three research assistants were joined on their first visit, three were joined on their second visit, and the remaining two were joined later in their first week; at post-test, six research assistants were joined on their first visit and six were joined on their second visit). No significant issues were identified. Some children were absent on the days or times when research assistants were scheduled to visit the schools and conduct the post-tests. Where feasible, research assistants made a second visit to try to assess all randomised pupils with priority given to Cohort 1 (i.e. children whose data would inform the primary outcome analysis).

Impact of attrition

In total, 91/93 (97.8%) of all randomised schools were retained within the evaluation; one school from the intervention and one from the control group withdrew from the evaluation.

While 9/47 (19.1%) schools randomised to the intervention group withdrew from delivering the TEEMUP PD, eight out of nine agreed to remain in the evaluation for post-testing.

Of the 1,566 Cohort 1 pupils who completed the baseline assessment and were included in the randomised sample, valid primary outcome post-test data were obtained for 1,361 (86.9%) pupils and 1,330 were subsequently included in the primary analysis (31 did not have valid EAL covariate data). Total pupil-level attrition between randomisation and the primary analysis sample was 236/1,566 (15.1%). Cohort 1 pupil-level attrition was equal across both the intervention (116/780, 14.9%) and control groups (120/786, 15.3%). Reasons for such pupil-level attrition are included in Figure 2.

Baseline characteristics for the Cohort 1 pupils as randomised and as included in the primary analysis are presented in Table 13 and Table 14 for visual comparison. No notable differences were observed.

Summary of key findings

- The delivery of the impact evaluation largely adhered to the trial protocol (Ainsworth *et al.,* 2023).
- A fifth of intervention schools formally withdrew from delivering the intervention (9/47), though all but one was retained for the evaluation and accommodated post-testing for the primary outcome.

- The impact evaluation experienced minimal attrition. Of the 93 randomised schools, 91 schools were retained within the evaluation for the primary outcome and child-level attrition from the randomised sample was low (at 15.1%) and matched what was assumed in the sample size calculation. Therefore, we deem the risk of sample attrition effects to be low.

Cost

For the purposes of the trial, schools allocated to receive the TEEMUP PD were not required to pay for training. Here we report the actual costs associated with implementing the programme, as they would be outside of this trial.

We estimated the average cost per pupil per year for schools receiving the TEEMUP PD programme following the EEF costing guidance issued in 2023 (EEF, 2023). A year is defined as a year of implementation. This may not align with the calendar or academic year. This costing model estimated costs based on the mean number of eligible pupils per school randomised into the evaluation (n = 50), since this is a whole-class intervention that covers both Reception and Year 1. This is in line with the average class size for infant classes (currently 26.6 children[7]); therefore, future interventions may be assumed to affect, on average, approximately 53 children per year.

Table 45 details the resources needed to implement the programme as per the ingredients method (Levin *et al.,* 2018). The main cost for implementing the programme is the cost of training. Other costs identified include staff cover, staff travel, and purchase of additional materials; however, these are categorised as optional rather than mandatory as only a proportion of schools utilised these.

Table 45: List of resources, 'ingredients'

Category		Item
Personnel for preparation and delivery		Schools – nominated Reception teacher(s)
		Schools – nominated Year 1 teacher(s)
		Schools – member of SMT to be lead contact
		Mentors
Personnel for training	Trainers	Oxford PD team
	Trainees	Schools – nominated Reception teacher(s)
		Schools – nominated Year 1 teacher(s)
		Schools – member of SMT to be lead contact (could attend workshops 1 and 2)
		Mentors
	Project management and admin	Oxford PD team
Facilities, equipment, and materials (prerequisites)		Information and communications technology equipment, e.g. laptop, computer, or iPad™, and internet connection
		Learning resources to support mathematical learning
Optional extras*		Additional learning resources to support mathematical learning (as needed)
		Staff cover (as needed)

**Some schools may need to purchase additional resources if they do not already have them in place as part of their usual provisions. Some schools may need to provide staff cover to release nominated teachers for training and intervention activities.*

Prerequisites

To take part in the TEEMUP PD programme, schools needed to have access to information and communications technology equipment such as a laptop or PC and an internet connection in order to access the dedicated TEEMUP website, which provided advice and resources. To our knowledge, no school purchased such equipment specifically for the purpose of taking part in the evaluation and we anticipate that most schools would already have this equipment should the programme be rolled out.

[7] https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics (accessed 15 August 2024).

While implementing the TEEMUP programme, schools were expected to continue with their usual CPD activities for teachers. A survey was conducted at the outset of the evaluation to understand what existing CPD programmes they were using and/or planned to use. Schools reported if they were signed up to take part of the NCETM Maths Hubs Programme, the White Rose Maths PD training, and other initiatives/CPD for EYFS/Key Stage 1 teachers in the coming academic year.

Time

The TEEMUP PD programme is intended to be delivered to nominated Reception and Year 1 teachers but can then be cascaded by them to other members of staff at the school. Given this, the time costs are largely front loaded and associated with completing training sessions, and schools potentially having to provide cover for teachers to do this, and continued development activities undertaken during the course of the implementation period (17 months in this evaluation). Training was provided via a combination of face-to-face workshops, live online sessions, or watching a previous, recorded live session. Some training took place in the second year of the trial to accommodate new teachers.

Data relating to the number and mode of training sessions (for the first two full days [six and a half hours each] and seven half-days [three hours and forty-five minutes]) attended by each teacher was provided by the delivery team and indicated that 95 teachers from the 38 intervention schools that did not withdraw from the intervention were nominated by their school to complete workshops. This is an average of 2.5 teachers per school. These teachers completed an average of 6.8 sessions (range 0 to 9). (NB. This is different to the figure presented for the CACE analysis as this was aggregated to the school level so data for only one Reception and one Year 1 teacher per school was used, whereas here we have considered all teachers.) Workshops were initially intended to be completed face-to-face; however, this was adapted so that sessions could be completed either face-to-face, or alternatively teachers could attend a trainer-led online training session, a trainer-led Zoom 'new teacher' session in Autumn Term 2022, or complete self-reported independent catch up.

For teachers involved in Cohort 1, the assumption was made that all sessions were completed in 2021–2022 unless any were indicated as a 'new teacher' session in Autumn Term 2022 in which case this was assumed to take place in 2022–2023 (this might be the case if a new teacher took over teaching Year 1 in 2022–2023). For teachers only involved in Cohort 2, it was assumed that all sessions were completed in 2022–2023.

In summary, 81 teachers from 38 schools completed sessions in 2021–2022; an average of 2.1 teachers per school, and 31.02 (SD 13.17) hours per teacher (Table 46). Twenty teachers from 17 schools completed sessions in 2022–2023; an average of 1.2 teachers per school (based on 17 schools, and 0.5 teachers per school across all 38 schools), and 15.79 (SD 15.48) hours per teacher (Table 46).

Surveys were distributed to nominated Reception and Year 1 teachers mid-implementation (completed by 39 teachers from 28 schools) and again at the end of the implementation period (completed by 54 teachers from 34 schools).

Teachers were asked whether their school arranged for paid cover for them to attend any/all of the TEEMUP PD workshops either at a venue, online, or independently, and if so to estimate the total number of hours of cover the school arranged. These data are summarised in Table 46. Twenty teachers (51.3% of 39) from 16 schools (57.1% of 28) said their school had paid for cover for them to complete workshops across the 2021–2022 academic year; the number of hours averaged 12.22 per teacher across all 28 schools (range 0 to 53). Thirty-two teachers (59.3% of 54) from 22 schools (64.7% of 34) said their school had paid for cover for them to complete workshops in the academic year 2022–2023; the number of hours averaged 9.70 per teacher across all 34 schools (range 0 to 53).

Additional time and financial resources were required for teachers to meet with their TEEMUP mentor, either online, over the phone, via email, or in person at the school. Via the surveys, teachers were asked to estimate, to the nearest hour, how much contact they had with their mentor, and how much paid cover was arranged during these times (Table 46). Thirteen teachers (33.3% of 39) from 11 schools (39.3% of 28) said their school had paid for cover for them to meet with their mentor over the academic year 2021–2022; the number of hours averaged 0.78 per teacher across all 28 schools (range 0 to 3). Twenty-two teachers (40.7% of 54) from 14 schools (41.2% of 34) said their school had paid for cover for them to meet with

their mentor in the academic year 2022–2023; the number of hours averaged 2.31 per teacher across all 34 schools (range 0 to 23).

Teachers were expected to spend time, outside of the workshops and mentor visits, completing activities relating to TEEMUP such as drafting/reflecting on a change plan, accessing the TEEMUP website, etc. They were asked to estimate how many hours in total they had spent on such activities, and the number of hours of paid cover the school arranged (Table 46). Three teachers (7.7% of 39) from three schools (10.7% of 28) said their school had paid for cover for them to complete other activities relating to the TEEMUP PD intervention in the academic year 2021–2022; the number of hours averaged 1.59 per teacher across all 28 schools (range 0 to 40). Five teachers (9.3% of 54) from three schools (8.8% of 34) said their school had paid for cover for them to complete other activities relating to the TEEMUP PD intervention in the academic year 2022–2023; the number of hours averaged 1.23 per teacher across all 34 schools (range 0 to 20).

Given the variable uptake of staff cover, this is considered an optional cost that would be utilised on a school-by-school basis.

Teachers were also asked to estimate the number of hours they spent on particular activities relating to the TEEMUP PD outside of their normal working school hours or in their own time (Table 47). Across the academic year 2021–2022, 27 teachers (69.2% of 39) from 23 schools (82.1% of 28) said that they completed TEEMUP PD workshops and/or related activities outside of normal working school hours/in their own time. Across the academic year 2022–2023, 20 teachers (37.0% of 54) from 17 schools (50.0% of 34) said that they completed TEEMUP PD workshops and/or related activities outside of normal working school hours/in their own time. Attending face-to-face training was cited as taking the most time.

Table 46: Total time devoted by personnel for training and related activities and staff cover

	Year 1 (2021–2022)	Year 2 (2022–2023)
	Mean (SD) no. of hours per teacher	Mean (SD) no. of hours per teacher
Time spent		
To complete workshops	31.02 (13.17)[a]	15.79 (15.48)[b]
To meet mentor	3.46 (1.73)[c]	6.87 (4.63)[d]
To complete other activities relating to TEEMUP	9.67 (9.27)[c]	18.94 (17.49)[d]
Hours of paid cover for teacher		
To complete workshops	12.22 (14.72)[c]	9.70 (13.8)[d]
To meet mentor	0.78 (1.22)[c]	2.31 (4.20)[d]
To complete other activities relating to TEEMUP	1.59 (7.08)[c]	1.23 (4.08)[d]

[a]*A total of 81 teachers from 38 schools.*
[b]*A total of 20 teachers from 17 schools.*
[c]*A total of 39 teachers from 28 schools.*
[d]*A total of 54 teachers from 34 schools.*

Table 47: Total time devoted by personnel for training and related activities outside working hours

	Year 1 (2021–2022)	Year 2 (2022–2023)
	Mean (SD) no. of hours per teacher[a]	Mean (SD) no. of hours per teacher[b]
Attending TEEMUP PD workshops at a venue	5.55 (11.23)	1.02 (4.50)
Attending TEEMUP PD workshops online	0.65 (1.28)	0.52 (1.79)
Completing TEEMUP PD workshops independently	0.87 (1.98)	0.59 (2.81)
Time spent with TEEMUP mentor	0.44 (1.17)	0.81 (2.05)
Completing other TEEMUP PD-related activities	3.33 (4.65)	3.56 (6.95)

[a]*A total of 39 teachers from 28 schools.*
[b]*A total of 54 teachers from 34 schools.*

Financial costs

Training (cost to Oxford)

Training of school staff took place either face-to-face or remotely. While the remote options were introduced in response to the COVID-19 pandemic, a similar model could be used to roll out the programme, based on school/teacher preference, capacity and availability, and ability of schools to provide cover for teachers to attend workshops.

Training of school staff was delivered by colleagues from the Oxford team. Teachers were trained in groups. Table 48 and Table 49 present the breakdown of programme delivery costs from the perspective of the Oxford team, these include intervention set-up and development, setting recruitment, and intervention delivery. These costs were provided by Oxford to the evaluation team and were broken down by whether or not they included additional COVID-19 overhead costs, which may not have been incurred if the trial had not been delivered during the pandemic.

These costs do not include the incentive payments made to schools to thank them for their participation in the trial as this was considered an evaluation cost that would not be incurred if the programme was rolled out.

Training (cost to schools)

Teachers were also asked to estimate how much it cost for them to travel to the venue(s) to attend the face-to-face TEEMUP PD workshops. The average was £49.05 (SD 67.38) in 2021–2022 (based on data from 39 teachers from 28 schools) and £37.91 (SD 67.92) in 2022–2023 (based on data from 54 teachers from 34 schools). This equates to £68.32 per school in 2021–2022 (Table 50) and £60.21 per school in 2022–2023. It should be noted that the original intention of the trial was to run training sessions in schools, reducing teacher travel time and costs, but due to the eventual randomisation timeline (randomisation was done in December with training starting in January), it was not possible to do this so external venues with a greater travel distance were used.

Staff cover

As detailed above, paid staff cover was used variably by schools and so it is considered an optional cost that schools considering implementing the TEEMUP PD programme should factor into their considerations. On average, across all schools providing data, cover was required for 14.59 hours per teacher (20.33 hours per school) in the academic year 2021–2022 and 13.24 hours per teacher (21.03 hours per school) in the academic year 2022–2023. To calculate the average cost per school, we multiplied these figures by £18.21. Therefore, on average, per school, £370.20 was spent on cover during the academic year 2021–2022 (Table 50) and £382.89 during the academic year 2022–2023.

Materials

Additional materials

In addition to being asked about their utilisation of existing materials, the survey asked if teachers had needed to buy any additional resources as a result of the TEEMUP PD in their school, and if so, what the item/resource was, the quantity, and the cost.

Thirteen teachers (33.3% of 39) from 11 schools (39.3% of 28) in the mid-implementation survey said they had needed to buy additional resources up to that date, of which ten teachers from eight schools provided details of the items bought, the quantity and cost. On average, teachers reported spending £56.97 on additional resources (range £0 to £595) in the year 2021–2022 (average £79.36 per school) (Table 50).

Twelve teachers (22.2% of 54) from 11 schools (32.4% of 34) in the end of trial survey said they had needed to buy additional resources up to that date, of which 11 teachers from ten schools provided details of the items bought, the quantity and cost. On average, teachers reported spending £23.02 on additional resources (range £0 to £300) in the year 2022–2023 (average £36.56 per school).

The most frequently cited items purchased were tens frames, rekenreks counting frames, books, and counters, other items included dice, games, and number/playing cards.

Since not all schools reported purchasing additional materials this is considered an optional rather than mandatory cost for implementation. Costs associated with additional resources could represent both a start-up or recurring cost depending on the frequency that resources need to be replaced. Costs are likely to be lower at start-up for schools that already have a wealth of relevant learning resources and higher for those that do not.

Overall costs

In total, Oxford estimates that the total cost to deliver the programme outside of an RCT in non-pandemic conditions for the 47 intervention schools amounted would be £545,948.44, which equates to £11,615.92 per school (Table 49).

Table 51 presents the total cost per pupil per year over three years based on the trial costs—£80.89. As stated above, this calculation assumes an average of 50 pupils per school and uses costs for delivering the programme outside of an RCT and excluding any COVID-19-specific costs.

The estimated cost of TEEMUP including COVID-19-specific overheads and costs, is £16,538 per school over a three-year period, or £110 per child per year when averaged over three years (Table 52).

Table 48: Cost of the implementation of the EEF programme, per mandatory ingredient—with COVID overheads

Cost ingredient	Start-up or recurring?	Nominal values			
		£ Year 1	£ Year 2	£ Year 3	Total
Setting recruitment	Start-up	£132,900.55	£0	£0	£132,900.55
Intervention delivery[a]	Start-up	£620,087.80	£0	£0	£620,087.80
Total cost of programme delivery for all schools		**£752,988.35**	**£0**	**£0**	**£752,988.35**
Total cost of programme delivery per school (47 schools)					**£16,021.03**
Total cost per pupil school year (50 pupils per school over three years)					**£106.81**

[a]*Staffing costs, venue hire, materials/resources for schools, mentor and trainer travel, website maintenance, project management and admin.*

Table 49: Cost of the implementation of the programme outside of an RCT, per mandatory ingredient

Cost ingredient	Start-up or recurring?	Nominal values			
		£ Year 1	£ Year 2	£ Year 3	Total
Intervention delivery[a]	Start-up	£545,948.44	£0	£0	£545,948.44
Total cost of programme delivery for all schools		**£545,948.44**	**£0**	**£0**	**£545,948.44**
Total cost of programme delivery per school (47 schools)					**£11,615.92**
Total cost per pupil school year (50 pupils per school over three years)					**£77.44**

[a]*Staffing costs, venue hire, materials/resources for schools, mentor and trainer travel, website maintenance, project management and admin.*

Table 50: Cost of the implementation of the programme, per optional ingredient

Category	Cost ingredient	Start-up or recurring?	Nominal values			
			£ Year 1	£ Year 2	£ Year 3	Total
Staff cover (start-up) (optional)	Paid cover for teachers to complete workshops, meet mentors, and conduct other TEEMUP activities	Start-up	£370.20	£0	£0	£370.20
Staff travel (start-up) (optional)	Staff travel to attend face-to-face workshops	Start-up	£68.32	£0	£0	£68.32
Materials (start-up) (optional)	Additional resources as a result of the TEEMUP PD	Start-up	£79.36	£0	£0	£79.36
Total cost of optional items per school			£517.88	£0.00	£0.00	£517.88
Total cost per pupil (50 pupils per school)						£10.36
Total cost per pupil school year (50 pupils per school over three years)						£3.45

Table 51: Total combined costs of training and optional extras, without COVID-19 overheads

Item	Type of cost	Total cost per school over three years	Total cost per pupil per year over three years (50 pupils per school per year)
Programme delivery (Oxford, without COVID-19 overheads)	Start-up cost per school	£11,615.92	£77.44
Optional cover	Start-up cost per school	£370.20	£2.47
Optional staff travel	Start-up cost per school	£68.32	£0.46
Optional materials	Start-up cost per school	£79.36	£0.53
Total		£12,133.80	£80.89

Table 52: Total combined costs of training and optional extras, including COVID-19 overheads.

Item	Type of cost	Total cost per school over three years	Total cost per pupil per year over three years (50 pupils per school per year)
Programme delivery (Oxford, including COVID-19 overheads)	Start-up cost per school	£16,021.03	£106.81
Optional cover	Start-up cost per school	£370.20	£2.47
Optional staff travel	Start-up cost per school	£68.32	£0.46
Optional materials	Start-up cost per school	£79.36	£0.53
Total		£16,538.91	£110.27

Conclusion

Table 53: Key conclusions

Key conclusions	
1.	Pupils in TEEMUP schools made the equivalent of one additional month's progress in maths attainment at the end of Year 1, on average, compared to pupils in other schools. These results have a high-security rating.
2.	FSM-eligible pupils in TEEMUP schools made the equivalent of two additional months' progress in maths attainment compared to pupils eligible for FSM in other schools. These results may have a lower security than the overall findings because of the smaller number of pupils.
3.	Teachers observed marked improvements in self-regulation and PSED outcomes for pupils, although it is unclear whether these were sustained in the longer term.
4.	Findings indicate that the TEEMUP PD was well received by teachers, who perceived a change in their own teaching and confidence. Teachers also perceived improvements to pupils' confidence in maths as well as a positive impact on children's self-regulation and behaviour within the classroom.
5.	The delivery of the programme was impacted by the COVID-19 pandemic and key elements of the training were adapted as a result. Competing commitments, which were exacerbated by the pandemic, limited the extent to which teachers could attend workshops and follow requirements of the programme.

Impact evaluation and IPE integration

Below we detail the relevant items of the logic model (in italics) and discuss the evidence base from the impact evaluation and the IPE in support (or not) of each logic model item. Overall, the results of the impact evaluation and the IPE provide evidence to support many components of the logic model. The majority of outputs, enabling factors, and environmental factors supported most of the short-term outcomes as per the IPE, which translated into improvements of varying degrees in both maths and PSED for both cohorts.

Outputs

Teachers engage with: workshops, a minimum of 75% of the time; at least three in-school mentoring sessions; and the bespoke knowledge sharing website. Teachers lead and evidence changes in the classroom, using material ideas, proformas, etc. provided, and engage further with families sharing maths resources and activities.

The results of the IPE and impact evaluation provide mixed evidence to support these 'outputs' of the logic model. Non-compliance among the 47 schools randomised to the intervention group was relatively high, with a third not meeting minimal engagement levels (with nine formally withdrawing from delivering the intervention). The impact and pressure of the COVID-19 pandemic on schools very likely influenced intervention withdrawals, as the reasons cited by schools were factors that were directly under severe pressure due to the COVID-19 pandemic; for example, staff changes and absences, and time and financial constraints. However, detailed compliance data collected from the 38/47 schools that did not withdraw from delivering the intervention demonstrated relatively high levels of engagement in terms of attendance at training workshops, accommodating mentor visits, and use of the TEEMUP website. Over 70% of these schools were considered to be by their TEEMUP mentor to have evidenced 'good' or 'excellent' use of the TEEMUP PD materials within the classroom for both cohorts, despite teachers reporting within the IPE that they felt they lacked time to consider how best to use and implement the resources/tools to make changes within their classrooms. Teachers who engaged with the intervention generally perceived the TEEMUP PD as high quality and applicable to practice. There was evidence within the IPE that intervention schools reported more frequently using formative assessment and reflective practice, more frequently considered ways to improve pupils' transition between Reception and engage with families to improve the home learning environment in comparison to baseline data and the control group. The frequency of training workshops coupled with the cost to schools of cover and staff changes in school were found to be barriers to fidelity. However, the trial was delivered in the backdrop of the COVID-19 pandemic, which saw schools experience high levels of staff absenteeism and turnover, which impacted on some schools' perceived and/or actual ability to engage with the intervention.

Enabling factors

Management and leadership in schools supportive of study, collaborations, and changes in schools; enthusiastic and committed participants. Teachers engage with website. Control schools continue with usual practice, and do not engage with any substantial maths PD during the study. Reception and Year 1 teachers remain stable (in same classes) throughout the project.

The IPE highlighted that the key necessary condition for the successful implementation of TEEMUP PD was that other school staff, particularly the headteacher and/or maths lead, was supportive of the study and of changes within the school. Schools tended to access the TEEMUP website very frequently over the course of the intervention, well above a frequency considered to be minimally compliant. A high proportion of schools (70%) in both the control and intervention group reported completing Key Stage 1 maths-related CPD programmes (other than TEEMUP) throughout the study period. Despite this, it does not compromise the design of the study as both intervention and control schools reported participation in similar amounts of additional maths PD. Reception teachers largely remained stable throughout the project. For Cohort 1, of the 38 intervention schools that did not formally withdraw from delivering the intervention, two reported having a change in Reception teacher part-way through the academic year 2021–2022 and four reported a change in teacher (or a job share) in Year 1 during the academic year 2022–2023. An estimated two-thirds of randomised pupils were moved to a Year 1 class being taught by a trained TEEMUP teacher, though the actual figure may be higher or lower than this as only 17 intervention schools provided this level of data. Three-quarters of the 38 schools retained a trained Reception teacher from 2021–2022 (Cohort 1) to teach Cohort 2 for the majority of the year in Reception in 2022–2023.

External moderators for all schools

School characteristics, for example, intake, deprivation levels, Ofsted rating, and size of school. Teacher characteristics, for example, length of time teaching. Pupil characteristics, for example, pupil deprivation, pupil EAL, and SEND status. Environmental risks, for example, school closures.

The majority of schools (87%) were rated 'Good' or 'Outstanding' on their latest Ofsted, and over half (~55%) reported partaking in the NCETM programme at some point during the evaluation. Our recruited sample of schools was roughly representative of the national picture in terms of deprivation as measured using the proxy of percentage of pupils currently eligible for FSM (18.8% of pupils in Reception year of state-funded primary schools, nationally, in 2021–2022 compared with 19.4% in our population). Participating teachers who completed the confidence surveys tended to have been a qualified teacher for approximately ten years, but there was quite a range, including newly or recently qualified teachers.

As specified in the eligibility criteria, pupils with significant SEND or EAL, which would prevent them from accessing the assessment and/or cause distress through completing the assessment, were excluded from the evaluation but those in participating intervention schools would still have experienced the intervention. In interviews, high levels of SEND within a classroom were identified by a minority of schools as a barrier to being able to implement TEEMUP PD resources aimed at improving PSED/self-regulation.

In subgroup analyses, Cohort 1 pupils eligible for FSM from schools that were offered TEEMUP made, on average, two additional months' progress in maths attainment at the end of Year 1 compared to pupils in the control group eligible for FSM. There was no evidence that the TEEMUP intervention was more or less effective for maths attainment at the end of Year 1 according to pupil gender or by whether the school was taking part in the NCETM Maths Hubs Programme for Cohort 1. For Cohort 2, the intervention appears to have been beneficial to boys but have a negative impact on maths attainment at the end of Reception for girls (as measured by the BAS3 ENC).

This evaluation was delivered in the backdrop and aftermath of the COVID-19 pandemic, which saw unprecedented challenges for the education sector, and resulted in some significant changes to TEEMUP PD training and challenges in implementation. The delivery team had to begin to offer online training to schools/teachers who were not able to attend face-to-face due to illness, absence, or difficultly finding cover. Even for training workshops delivered at a venue, adaptations had to be made to minimise schools' collaborations as schools requested their teachers remained within 'bubbles'. This altered the intended teacher experience of the TEEMUP PD training session. Despite this, IPE respondents

commented on the preference of the face-to-face training over online, as it offered an opportunity for collaboration, albeit a more limited one than intended.

Short-term outcomes/mediators

Teachers show increased knowledge and understanding of maths and self-regulation (confidence) and how to teach it (change in classroom practice). Staff planning and interactions are more responsive to children's needs. Teachers identify children in need of more maths and self-regulation support and target them more effectively. Teachers report pupil engagement with maths increased. Teachers work more collaboratively with staff in class and across Reception and Year 1. Teachers report improved partnership working with families around the home learning environment. Teachers report new practices embedded in the second year of PD.

Teachers in the intervention group reported greater confidence than those in the control group when asked both mid- and post-implementation of the intervention.

Among schools that did not withdraw from intervention delivery, there was evidence that a high proportion of them had changed their classroom practice (as per mentor-determined 'evidence of change' criteria) as a result of the TEEMUP PD. There was also quite good evidence in the IPE of teacher-reported improvements to planning, interactions with children, and identifying children in need, or more maths and self-regulation support, and that teachers planned to embed them into their practice from the first to second year of intervention delivery, and beyond the context of trial. There was strong evidence within the IPE that pupils' engagement with maths had increased. A short-term outcome/mediator within the original logic model (Figure 1) was that teachers work more collaboratively with staff in class and across Reception and Year 1. Interview data from schools would suggest there is some evidence that teachers working more collaboratively across Reception and Year 1 occurred in some of the schools that participated in interviews.

Long-term outputs/impacts of TEEMUP

Improved maths attainment at the end of Reception and Year 1. Improved PSED and self-regulation at the end of Reception and Year 1. Improved attainment in maths and self-regulation for children from disadvantaged backgrounds. Improved staff understanding and implementation of developmentally appropriate practice in particular with maths instruction (self-report). Improved partnership working with families on maths home learning environment (self-report).

Small improvements to maths attainment—the equivalent of one month's additional progress—were observed among Cohort 1 following five to six months of the intervention (a secondary outcome measured by teacher assessed EYFSP when Cohort 1 pupils were at the end of Reception) and following 17 months of the intervention (the primary outcome measured by the BAS3 ENC collected by independent, blinded research assistants when Cohort 1 pupils were at the end of Year 1). Disadvantage pupils also made gains in maths attainment—the equivalent of two months' additional progress—in comparison to the control group. Within the IPE, teachers reported mixed observations as to which pupils they considered benefited most from the TEEMUP PD; however, disadvantaged pupils were not identified as a specific theme. Teachers complying with the intervention increased maths attainment among Cohort 1 by two months. Though none of these results reached statistical significance.

There was no improvement to maths attainment at the end of Reception for Cohort 2 as measured by the BAS3 ENC (secondary outcome) but moderate gains—the equivalent of four months' additional progress—were observed when measured using the teacher assessed EYFSP (secondary outcome); indicating a positive, statistically significant effect of the TEEMUP PD on pupils' maths attainment in comparison to the control group. Reception teachers' compliance with the intervention did not increase Cohort 2's maths attainment. These results relating to maths attainment are somewhat concordant with the results of the IPE, where staff who were interviewed and surveyed perceived observable changes to pupils' maths development and understanding and, for some, improvements to pupils' maths attainment using their own assessment measures.

Small to large improvements to PSED and self-regulation—the equivalent of two to five months' additional progress—were observed among Cohort 1 following five to six months of the intervention (secondary outcome measured by teacher

assessed CSBQ when Cohort 1 pupils were at the end of Reception); indicating a positive, statistically significant effect of the TEEMUP PD on pupils cognitive, emotional, and behavioural self-regulation, sociability, and externalising and internalising problems in comparison to the control group. When measured using the teacher assessed EYFSP (secondary outcome), PSED improved by one month, although this result was not significant. The gains observed among Cohort 1 at the end of Reception were not sustained at the end of Year 1 following 17 months of the intervention, with no improvements to PSED and self-regulation reported in the CSBQ data (secondary outcome, teacher assessed); in fact, the results mostly favoured the control group.

Small to moderate improvements to PSED and self-regulation—the equivalent of two to four months' additional progress—were observed among Cohort 2 at the end of Reception when measured using teacher assessed outcomes. These data indicate a positive, significant effect of the TEEMUP PD—on pupils cognitive, emotional, behavioural self-regulation, sociability, and prosocial behaviour, and internalising problems domains, as measured by the CSBQ—in comparison to the control group. EYFSP data show a two-month improvement in pupils' mean ELG score (the mean point score for the 17 ELG), also a statistically significant result.

There was evidence within the IPE that schools had improved partnership working with families on maths home learning environment. The findings from the impact and the IPE together were deducted from a cohort of intervention schools where non-compliance/intervention delivery withdrawal was relatively high.

Interpretation

This evaluation was delivered in the backdrop and aftermath of the COVID-19 pandemic, which impacted both the delivery of this intervention and the collection of data for the evaluation. The pandemic saw unprecedented challenges for the education sector. At a pupil level, Ofsted (2021) reported that the implications of the pandemic on the 2021 Reception intake (Cohort 1 within this evaluation) was bigger than expected—in part due to missing out on or having interrupted nursery provision—resulting in pupils having a wider range of starting points and struggling more with peer interactions, behaviour, school readiness, and attitudes to learning. For many schools in England, pupil-level attendance between 2021 and 2023—the entire period of the trial—had not returned to pre-pandemic levels (Long and Roberts, 2024), with schools reporting more COVID-19-related absences among disadvantaged pupils and those with SEND. At a teacher level, schools experienced higher teacher sickness absence during 2021–2022 in comparison to pre-pandemic records (DfE, 2023c), often with multiple staff members absent at the same time (Ofsted, 2021). Between 2021 and 2023, schools also experienced high staff turnover, with double the number of school staff vacancies in comparison to before the pandemic (DfE, 2023c). It is important that this context is considered when interpreting the results of this evaluation.

Mathematical attainment later in life can be predicted by mathematical achievement early in a child's school career (Duncan *et al.,* 2007; Jordan *et al.,* 2009; Claessens, Engel, and Curran, 2014; Nguyen *et al.,* 2016). It is recommended that CPD should be implemented in schools to improve the quality of Key Stage 1 teachers' knowledge of maths, of pupils' mathematical development, and of effective mathematical pedagogy (Clark, Henderson, and Gifford, 2020). The TEEMUP PD evaluated within this efficacy trial was a modified version of a CPD programme delivered to early years settings of four–five-year-olds in Australia, referred to as the FEEL study (Siraj *et al.,* 2018a). The TEEMUP PD programme was designed to develop mathematics teaching among teachers of pupils in Reception (four–five-year-olds) and Year 1 (five–six-year-olds) to reduce the proportion of pupils who do not meet the expected standard in maths, and to improve PSED and self-regulation among these age groups. The TEEMUP PD intervention was an intensive CPD intervention, offering up to 53 hours of face-to-face workshops, ongoing communication and meetings with a mentor, as well as access to tools and resources for teachers to implement within their practice. As detailed in the TIDieR table (Table 3), the delivery of some components of the TEEMUP PD underwent adaptations in response to the COVID-19 pandemic, for example, providing training workshops online as opposed to only face-to-face and changes to the format of face-to-face workshops to allow for social distancing.

The evaluation took the form of a large (93 schools), two-armed, cluster RCT aiming primarily to improve mathematics among Reception and Year 1 pupils. Randomisation used minimisation and the school and pupil characteristics were well balanced between trial arms at baseline. School-level attrition, a potential source of post-randomisation bias, was low in this trial, with only 2/93 (one control, one intervention) randomised schools fully withdrawing from the evaluation. Post-

randomisation pupil-level attrition was also relatively low (particularly for a two-year trial) at 15% and was evenly balanced between the two groups. Characteristics for the 1,330 pupils included in the primary analysis were very similar to the 'as randomised' population and, therefore, there was no evidence that this attrition had introduced selection bias or any baseline imbalances among pupils that were included in the primary analysis. Recruitment to Cohort 2 (collecting secondary outcome data only) was affected by the withdrawal of 18 schools that declined to recruit a new Reception cohort in the second year of the trial (academic year 2022–2023). Outcome assessment (BAS3 ENC) was undertaken by independent research assistants who were trained by the evaluation team and 'blind' to the allocation of the pupils being tested. The statistical power of the study was higher than expected as the ICC was lower than assumed (0.04 as opposed to 0.15), which meant that the MDES was 0.17 rather than 0.21 as stated in the sample size calculation.

Generally, the intervention was very well received by teachers in schools delivering the intervention, with compliance data indicating good adherence to most components of the TEEMUP PD. The results of the primary outcome—pupils mathematics attainment at the end of Year 1 after two years of intervention delivery (Cohort 1)—show a small positive effect size of 0.07 for the intervention group that was not statistically significant (95% CI: -0.05 to 0.18). However, the CI includes the possibility that the intervention could have no impact or a larger impact of up to two additional months' progress but still excluded the original MDES difference sought in the Statistical Analysis Plan of 0.21. The primary maths attainment outcome finding is therefore, concordant with the results of the IPE, where staff perceived observable changes to pupils' maths development and understanding, and for some, improvements to pupils' maths attainment using their assessment measures.

There was an element of non-compliance with 9/47 schools (19.1%) randomised to the intervention formally withdrawing from the intervention; although eight out of nine of these schools were retained in the evaluation and provided post-test primary outcome data, which allowed an ITT analysis to be undertaken. High intervention withdrawal rates could dilute any treatment effect. Indeed, a CACE analysis exploring the effects of non-adherence on the primary outcome suggested that increased compliance led to greater improvements; among schools that were deemed to be good compliers, the effect size was moderate and favoured the intervention group (0.12).

Pupils in intervention schools in Cohort 1 made increases in the self-regulation (CSBQ) subscales (which were statistically significant except for prosocial) at the end of Reception; however, the gains were not maintained for Cohort 1 pupils at the end of Year 1 when the effect sizes favoured the control groups for five of the eight domains assessed.

Outcomes obtained from the EYFSP at the end of Reception indicated a small positive difference in maths attainment between the intervention and control groups for Cohort 1.

For Cohort 2 pupils, whose mathematics attainment was assessed at the end of Reception, the results showed no effect of the intervention (effect size of 0.01, 95% CI: -0.14 to 0.16). These results are in contrast to the findings reported by the FEEL study, which observed small positive improvements to early numeracy once controlling for baseline imbalances among four–five-year-olds, although the improvement did not meet statistical significance (Siraj *et al.,* 2018a). We did not have a baseline assessment for the BAS3 ENC in Cohort 2, which likely resulted in some loss in power for the outcome analysis; while the correlation between baseline and post-test BAS3 ENC for Cohort 1 was only moderate at 0.47, this was over two years rather than over one year as it would have been for Cohort 2. The Cohort 2 analysis did adjust for a lagged, school-level mean BAS3 ENC score from Cohort 1, which only loosely correlated with the outcome (0.11) and so there was minimal difference in the results when this covariate was omitted in sensitivity analyses.

Despite relatively high school-level intervention withdrawal rates as described earlier in this section, a CACE analysis, exploring the effects of non-adherence on Cohort 2's maths attainment did not find that increased compliance lead to greater improvements.

Pupils in intervention schools in Cohort 2 made increases in the self-regulation (CSBQ) subscales (which were statistically significant except for externalising) at the end of Reception. Outcomes obtained from the EYFSP at the end of Reception for Cohort 2 indicated statistically significant improvements in maths, as measured by the Mathematics ELGs, and in the average score among the 17 ELGs assessed.

A number of subgroup analyses were undertaken. Cohort 1 FSM-eligible pupils in TEEMUP schools made the equivalent of two additional months' progress in maths attainment compared to pupils eligible for FSM in other schools. These results may have lower security than the overall findings because of the smaller number of pupils. Further subgroup analyses did indicate that boys benefited from the intervention in Cohort 2, while it had a negative impact on girls. However, generally, there was no consensus within the IPE data as to whether teachers perceived the intervention to have benefited some groups of children more than others.

Participating schools/teachers were advised to retain nominated Reception and Year 1 teachers in their respective year groups for the duration of the trial so that Reception teachers received training during 2021–2022 preparing them to fully implement the PD into teaching practice in the academic year 2022–2023. Indeed, most schools retained a trained Reception teacher from 2021–2022 (Cohort 1) to teach Cohort 2 for the majority of the year in Reception in 2022–2023 (it is common for Reception teachers to remain teaching Reception year and not move to other year groups). However, the impact of this 'soak time' is not borne out in the results relating to maths attainment as measured by the BAS3 ENC for Cohort 2, though there is evidence for this in maths attainment measured by the EYFSP.

Notable improvements in self-regulation/PSED were observed across both cohorts at the end of Reception, with the gains tending to be slightly greater for Cohort 1. Such gains could be a reflection of Cohort 1's lower levels of school readiness and PSED skills due to the COVID-19 pandemic, as reported by Ofsted (Ofsted, 2021). As detailed within the EEF's Early Learning Toolkit relating to self-regulation strategies, studies indicate that the most notable improvements in the areas of self-regulation PSED often occur during the early years. This is because children are particularly receptive to learning and development in these areas during this time. This makes the early years a critical window for interventions aimed at impacting these areas. Early gains in PSED/self-regulation can have a lasting positive impact on a child's well-being and attainment. However, the gains to self-regulation/PSED observed in Reception within this trial were largely not sustained at the end of Year 1 (Cohort 1). This could be due to a number of factors including the high levels of intervention withdrawal observed within the trial, and/or the fact the TEEMUP training offered to and received by some teachers varied in format and intensity to what was initially intended (mainly due to the impact COVID had on schools and the delivery of the intervention, discussed further below). Additionally, approximately one-third of Cohort 1 pupils did not move into a Year 1 class taught by teachers at least 'minimally' trained in TEEMUP; which may have diluted the treatment effect. At recruitment, schools were requested to commit to moving participating Cohort 1 pupils from a TEEMUP trained Reception teacher to a trained Year 1. However, this was evidently difficult for some schools to achieve, likely in part due to the unprecedented staff turnover rates within schools during 2021–2023 due to the COVID-19 pandemic (ONS, 2023).

Such high levels of staff turnover were not anticipated at the development stages of the intervention or evaluation design, which both took place pre-pandemic. Nevertheless, these findings suggest it is not always possible for teachers to remain in the same class from year to year, or pupils to remain in the same class as they move to different year groups. This has implications for the sustainability of interventions and initiatives that are tailored specifically for use in certain year groups and the evaluation of them within research.

Schaeffer *et al.* (2021) found that pupil learning can be affected by teacher's confidence in teaching maths. McCullouch (2016) suggested that primary teachers 'lack of confidence in teaching maths may arise from weak understanding of the subject, limiting the teacher's ability to extend and enhance the children's progress' (McCulloch, 2016; p. 9). The impact evaluation found evidence that the TEEMUP PD improved teachers' confidence in their maths teaching practice and a dominant view within the IPE data was that staff perceived this to be the case.

A high proportion (70%) of all participating schools were also engaging with other substantial maths-related CPD in the first year of the study. Similarly, Ofsted (2021) reported that most primary schools were also focusing on catch-up work in mathematics during 2021–2022 in response to the COVID-19 pandemic, which may help to explain the high proportion of maths CPD undertaken by schools participating in the trial. Teachers who engaged with the intervention perceived the TEEMUP PD as high quality and applicable to practice; however, the intensity of the PD coupled with high staff absence was identified as a barrier to fidelity. Teachers also lacked focused time to consider how best to implement TEEMUP PD within their classrooms. Whether these barriers to fidelity and implementation would have been as dominantly reported pre-pandemic is unknown. However, the time commitment required from teachers to participate in the TEEMUP PD (48 hours

of workshops in the first year, plus time with the school's mentor and time to engage with resources) is substantially higher than the average 21.1 hours teachers in the study reported usually being afforded across an academic year for CPD.

As with most evaluations, the results of the study are only applicable to the sample that were enrolled into the study. The proportion of schools recruited to the study out of those approached (approximately 4%, i.e. 93/2,356) was relatively small, which could introduce selection bias in that the schools enrolled into the study were not representative of the schools approached for the trial or of the 'average' state primary school. School recruitment for the trial took place during the peak of the COVID-19 pandemic, when schools in England faced multiple, prolonged closures; this context may have negatively affected school-level recruitment at the time. When school- and pupil-level characteristics of the randomised sample were compared with national data we saw some small differences. We had similar levels of pupils eligible for FSM but had fewer pupils with EAL, perhaps because children with EAL where an extreme language barrier existed or who were new to English were excluded as this would prevent them from accessing the assessment and/or cause distress through completing the assessment. In our sample, relative to the national picture, we had a smaller proportion of community schools and a slightly higher proportion of academy converter schools; a smaller proportion of schools in urban areas, as opposed to rural areas; and, of the schools for which the latest Ofsted inspection rating was available, we have a lower proportion of schools rated as 'Outstanding' (9.1%) and a higher proportion of schools rated as 'Good' (83.0%). Therefore, the results may not apply to schools that were not eligible to take part in the study.

Limitations and lessons learned

The trial recruited reasonably well and there was low attrition to the primary outcome. However, response rates to the teacher-reported CSBQ were low in Cohort 1 (approximately 60% and 70% at the midpoint and endpoint, respectively). Nearly a fifth of randomised schools declined to recruit to Cohort 2; however, those that did remained relatively engaged with a 90% response rate to the CSBQ at post-test. There are limitations to some of the analyses conducted and variables used, including how participation by a school in the NCETM Maths Hubs Programme was collected and defined. The NCETM Maths Hubs Programme was a concurrent intervention, which also aimed to improve maths' attainment, that schools could be involved in; therefore, it was important to consider whether this impacted TEEMUP trial results via subgroup and sensitivity analyses. For the purposes of these analyses, a school was considered to be taking part in the NCETM Maths Hubs Programme if a contact from the school completed at least one of the three surveys throughout the trial and responded that the school had signed up to the NCETM Maths Hubs Programme at any time point. However, some schools did not respond to any of the surveys, so it is possible there were schools that participated in the NCETM Maths Hubs Programme that were not identified. For those that were, no further data were collected to indicate the level of participation, for example, whether teachers joined face-to-face or online courses or merely joined hub meetings. Despite these limitations, the results when the primary analysis was repeated including NCETM participation as a covariate were virtually identical to the primary results, so it seems unlikely participation by schools in the Maths Hub impacted the findings from this trial. In addition, it was not possible to collect baseline BAS3 ENC scores for Cohort 2. While a pre-test is not a necessity in an RCT, it can help improve the power of the trial if it is well correlated with the outcome. As an alternative to pupil-level baseline data from Cohort 2 pupils, we used a lagged school-level measure of prior attainment by calculating the mean baseline BAS3 ENC score per school from Cohort 1. The correlation between this proxy pre-test measure and outcome for Cohort 2 proved to be very low; therefore, the results when including this in the analysis model for the BAS3 ENC for Cohort 2 were very similar to when this covariate was omitted. The results of the teacher confidence survey may also not be very reliable due to the low response rate over time as only 78 teachers provided data at baseline and one follow-up time point (n=78).

We acknowledge the possibility that schools that participated in IPE interviews may have been 'systematically' different from schools that did not volunteer to take part in these aspects and, consequently, the results should be treated with a measure of caution, as there is a possibility of a potential source of bias having been introduced due to the self-selection process. For example, teachers in schools, which were extremely enthusiastic about the TEEMUP PD may have been more likely to volunteer to be interviewed and, conversely, the less engaged schools or those with perceived workload issues may have been less likely to take part.

We acknowledge the potential for conflict of interest in the delivery team interview's data and, therefore, this is treated with a measure of caution. This has to be balanced with the fact that the delivery team provided a unique and valid perception of the programme and their insights supplement the data from the other stakeholder interviews.

The purpose of the longitudinal interviews was to follow schools for the entirety of the TEEMUP PD journey through four interviews and by observing one mentoring session. The four intended interview time points throughout intervention delivery offered the opportunity for stakeholders to chronologically capture their perception of the intervention. However, we experienced difficulties recruiting schools and retaining teachers' participation to this longitudinal component of the IPE (we underrecruited by one school and of the three schools who participated, one withdrew from delivering the TEEMUP PD, one experienced staff changes, and it was not possible to observe a mentoring session in the remaining school. Therefore the longitudinal interview dataset was not as expansive as originally planned due to attrition. To compensate, (1) additional schools were recruited to take part in interviews at two time points to explore intervention delivery across schools with varying levels of compliance, and (2) a question was added to the endpoint intervention survey about teachers' experiences with the TEEMUP mentor. . Recruitment for longitudinal interviews took place during the most intensive stage of intervention delivery where teachers needed to attend TEEMUP PD training workshops, which likely impacted on staff capacity and willingness to participate within additional evaluation activities. The report has highlighted the high levels of intervention withdrawal within the trial, and the issues schools experience in relation to staff retention, which were likely exacerbated by the COVID-19 pandemic.

In Autumn Term 2021, teacher-level usual practice surveys were administered pre-randomisation to Reception and Year 1 teachers who were nominated by their schools to receive the TEEMUP PD should they be allocated following randomisation. As noted within the report, there were a high proportion of staff changes between the start and end of the trial (in Summer Term 2023). We acknowledge that not every teacher will have completed both surveys, and some teachers will have completed the first but not the final. This is also the case for intervention delivery surveys issued in Summer Term 2022 and Summer Term 2023. This may result in response bias and is a limitation of the comparisons made between survey data.

Future research and publications

This efficacy trial was severely impacted by the COVID-19 pandemic. Despite this, small improvements to maths—particularly for disadvantaged pupils—and small, moderate improvements to PSED and self-regulation for all pupils were found. Given these findings, the intervention may benefit from a re-trial without the noise of a global pandemic. Consideration should be given as to the appropriateness of a two-year trial in which pupils move from a trained Reception teacher to a trained Year 1 teacher as, evidently, this may be an impractical scenario for many schools and something which cannot be guaranteed or enforced.

A number of studies suggest that improving the self-regulation skills of children in the early years is likely to have a lasting positive impact on later learning at school (e.g. Skibbe *et al.*, 2019). Given the positive impact of the TEEMUP PD on pupils, future research could explore the impact of the TEEMUP PD delivered by fully trained Reception teachers (i.e. teachers that are TEEMUP trained prior to the start of the academic year in which the cohort of interest start Reception) on the impact of children beyond the end of Reception. In such a scenario, schools with nurseries would be included, and a nursery pupil cohort—who intend to continue their education at the same school come Reception age—could be recruited and assessed at baseline in the autumn prior to starting Reception.

Future research could investigate if trained teachers continue to implement the TEEMUP PD and the impact that increased 'soak time' has on staff's teaching practice and, consequently, children's outcomes.

References

Ainsworth, H., Fairhurst, C., Hugill-Jones, J., Leggert, H., Whiteside, K., Baird, K., Elliott, L., Fountain, I., Bell, K., Torgerson, C., Torgerson, D. and Robinson-Smith, L. (2023) '*Independent Evaluation of the Oxford Teaching Effective Early Mathematics and Understanding in Primary schools (TEEMUP) Professional Development Programme: A Two-Armed Cluster Randomised Controlled Trial. Evaluation Protocol*'. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/teem-up (accessed 6 January 2023).

Baird, K. and Fairhurst, C. (nd) *'Independent Evaluation of the Oxford Teaching Effective Early Mathematics and Understanding in Primary schools (TEEMUP) Professional Development Programme: A Two-Armed Cluster Randomised Controlled Trial: Statistical Analysis Plan'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/teem-up(accessed 12 February 2024).

Bibby, T. (1999) 'Subject Knowledge, Personal History and Professional Change'. *Teacher Developments*, 3: 2, 219–232. https://doi.org/10.1080/13664539900200084

Bibby, T. (2002) 'Shame: An Emotional Response to Doing Mathematics as an Adult and a Teacher'. *British Educational Research Journal*, 28: 5, 705–721. https://doi.org/10.1080/0141192022000015543

Braun, V. and Clarke, V. (2006) 'Using Thematic Analysis in Psychology'. *Qualitative Research in Psychology*, 3: 2, 77–101. https://doi.org/10.1191/1478088706qp063oa

Brophy, J. (1986) 'Teacher Influences on Student Achievement'. *American Psychologist*, 41: 10, 1069–1077. https://doi.org/10.1037/0003-066X.41.10.1069

Bynner, J. and Parsons, S. (1997) *Does Numeracy Matter?* London: The Basic Skills Agency.

Bynner, J. and Parsons, S. (2000) 'Impact of Poor Numeracy on Employment and Career Progression'. In: C. Tikly and A. Wolf (Eds.), *The Maths we Need Now: Demands, Deficits and Remedies*. London: Institute of Education. pp.26–51.

Charalambous, C.Y. and Praetorius, A.-K. (2018) 'Studying Mathematics Instruction Through Different Lenses: Setting the Ground for Understanding Instructional Quality More Comprehensively'. *ZDM Mathematics Education*, 50: 3, 355–366. https://doi.org/10.1007/s11858-018-0914-8

Chen, J.-Q., McCray, J., Adams, M. and Leow, C. (2014) 'A Survey Study of Early Childhood Teachers' Beliefs and Confidence About Teaching Early Math'. *Early Childhood Education Journal*, 42: 6, 367–377. https://doi.org/10.1007/s10643-013-0619-0

Claessens, A. and Engel, M. (2013) 'How Important is Where You Start? Early Mathematics Knowledge and Later School Success'. *Teachers College Record*, 115: 6, 1–29. https://doi.org/10.1177/016146811311500603

Claessens, A., Engel, M. and Curran, F.C. (2014) 'Academic Content, Student Learning, and the Persistence of Preschool Effects'. *American Educational Research Journal*, 51: 2, 403–434. https://doi.org/10.3102/0002831213513634

Clark, A., Henderson, P. and Gifford, S. (2020) '*Improving Mathematics in the Early Years and Key Stage 1: Guidance Report'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/early-maths (accessed 15 February 2024).

Clements, D.H. and Sarama, J. (2011) 'Early Childhood Mathematics Intervention'. *Science*, 333: 6045, 968–970. https://doi.org/10.1126/science.1204537

Coe, R., Aloisi, C., Higgins, S. and Major, L.E. (2014) '*What Makes Great Teaching? Review of the Underpinning Research*'. p. 58. Available at: www.suttontrust.com/wp-content/uploads/2014/10/What-Makes-Great-Teaching-REPORT.pdf (accessed 27 April 2020).

Cook, T.D. and Campbell, D.T. (1979) '*Quasi-Experimentation: Design and Analysis Issues for Field Settings*'. Boston: Houghton Mifflin.

Cook, T.D., Campbell, D.T. and Shadish, W. (2002) '*Experimental and Quasi-Experimental Designs for Generalized Causal Inference*'. Boston, MA: Houghton Mifflin.

Council of the European Union. (2002) 'Council Resolution of 27 June 2002 on Lifelong Learning'. *Official Journal of the European Communities*, 2002/C(163/01). Available at: https://op.europa.eu/en/publication-detail/-/publication/0bf0f197-5b35-4a97-9612-19674583cb5b (accessed 12 February 2024).

Cuschieri, S. (2019) 'The CONSORT Statement'. *Saudi Journal of Anaesthesia*, 13: Suppl 1, S27–S30. https://doi.org/10.4103/sja.SJA_559_18

Data Protection Act 2018, c.12. Available at: www.legislation.gov.uk/ukpga/2018/12/contents (accessed 12 February 2024).Department for Education (DfE). (2019) '*National Curriculum Assessments at Key Stage 2 in England 2019 (Revised)*'. London: Department for Education. Available at: www.gov.uk/government/statistics/national-curriculum-assessments-key-stage-2-2019-revised (accessed 21 May 2021).

Department for Education (DfE). (2023a) '*Key Stage 2 Attainment: Academic Year 2022/23*'. GOV.UK. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-2-attainment/2022-23 (accessed 12 February 2024).

Department for Education (DfE). (2023b) '*Early Years Foundation Stage Profile: 2024 Handbook*'. London: Department for Education. Available at: www.gov.uk/government/publications/early-years-foundation-stage-profile-handbook (accessed 12 February 2024).

Department for Education (DfE). (2023c) '*School Workforce in England: Reporting Year 2022*'. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/school-workforce-in-england (accessed 15 February 2024).

Department for Education (DfE). (2024). Early years foundation stage statutory framework. For group and school-based providers. Available at: https://www.gov.uk/government/publications/early-years-foundation-stage-framework--2#full-publication-update-history

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K. and Japel, C. (2007) 'School Readiness and Later Achievement'. *Developmental Psychology*, 43: 6, 19. https://doi.org/10.1037/0012-1649.43.6.1428

Dunn, G., Maracy, M. and Tomenson, B. (2005) 'Estimating Treatment Effects From Randomized Clinical Trials With Noncompliance and Loss to Follow-Up: The Role of Instrumental Variable Methods'. *Statistical Methods in Medical Research*, 14: 4, 369–395. https://doi.org/10.1191/0962280205sm403oa

Easton, C. and Gee, G. (2012) '*Early Intervention: Informing Local Practice (LGA Research Report)*'. Slough: NFER. Available at: www.nfer.ac.uk/publications/LGLC02/LGLC02.pdf (accessed 8 October 2019).

Education Endowment Foundation (EEF). (2022a) '*Statistical Analysis Guidance for EEF Evaluations*'. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design (accessed 15 February 2024).

Education Endowment Foundation (EEF). (2022b) '*Implementation and Process Evaluation Guidance for EEF Evaluations*'. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design (accessed 15 February 2024).

Education Endowment Foundation (EEF). (2023) *'Cost Evaluation Guidance for EEF Evaluations'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design (accessed 12 February 2024).

Elliot, C.D. and Smith, P. (2011) '*British Ability Scales: Third Edition (BAS3)'*. London: GL Assessment.

Ellis, S. and Tod, J. (2015) '*Promoting Behaviour for Learning in the Classroom: Effective Strategies, Personal Style and Professionalism*. Abingdon, UK: Routledge.

Franke, M.L. and Kazemi, E. (2001) 'Learning to Teach Mathematics: Focus on Student Thinking'. *Theory Into Practice*, 40: 2, 102–109. https://doi.org/10.1207/s15430421tip4002_4

Ge M, Durham LK, Meyer RD, Xie W, Thomas N. (2011) 'Covariate-Adjusted Difference in Proportions from Clinical Trials Using Logistic Regression and Weighted Risk Differences'. *Drug Information Journal*. 45: 4, 481–493. https://doi.org/10.1177/009286151104500409

General Data Protection Regulation (GDPR). (2016) *'Council Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (United Kingdom General Data Protection Regulation) (Text with EEA relevance)'*. Available at: www.legislation.gov.uk/eur/2016/679 (accessed 12 February 2024).

Halle, T., Metz, A. and Martinez-Beck, I. (2013) '*Applying Implementation Science in Early Childhood Programs and Systems'*. p. 23. Baltimore: Brookes Publishing Company.

Hentges, R.F., Galla, B.M. and Wang, M. (2019) 'Economic Disadvantage and Math Achievement: The Significance of Perceived Cost From an Evolutionary Perspective'. *British Journal of Educational Psychology*, 89: 2, 343–358. https://doi.org/10.1111/bjep.12242

Howard, S.J. and Melhuish, E. (2017) 'An Early Years Toolbox for Assessing Early Executive Function, Language, Self-Regulation, and Social Development: Validity, Reliability, and Preliminary Norms'. *Journal of Psychoeducational Assessment*, 35: 3, 255–275. https://doi.org/10.1177/0734282916633009

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K. and Buck, R. (2016) *Implementation and Process Evaluation (IPE) for Interventions in Educational Settings: An Introductory Handbook'*. p. 52. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/IPE_Handbook.pdf (accessed 12 February 2024).

Jordan, N.C., Kaplan, D., Ramineni, C. and Locuniak, M.N. (2009) 'Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes'. *Developmental Psychology*, 45: 3, 850–867. https://doi.org/10.1037/a0014939

Jordan, N.C. and Levine, S.C. (2009) 'Socioeconomic Variation, Number Competence, and Mathematics Learning Difficulties in Young Children'. *Developmental Disabilities Research Reviews*, 15: 1, 60–68. https://doi.org/10.1002/ddrr.46

Kingston, D.J. (2017) '*A Mixed Methods Study: Evaluating the Impact of a Bespoke Professional Development Based on an Analysis of Existing Quality in One Local Authority'*. Doctoral Thesis. London: UCL. Available at: http://discovery.ucl.ac.uk/1551541/ (accessed 12 February 2024).

Knowles, C. and Fair Education Alliance. (2017) '*Closing the Attainment Gap in Maths: A Study of Good Practice in Early Years and Primary Settings'*. London: Fair Education Alliance. Available at: https://static1.squarespace.com/static/543e665de4b0fbb2b140b291/t/58aaeac429687f223f0ff369/1487596235907/FEA+Numeracy+Report_FV.pdf (accessed 21 May 2021).

Levin, H.M., McEwan, P.J., Belfield, C., Bowden, B. and Shand, R. (2018) *'Economic Evaluation in Education: Cost-Effectiveness and Benefit Cost Analysis, Third Edition'*. Thousand Oaks, USA: SAGE Publications, Inc.

Long, R. and Roberts, N. (2024) '*School Attendance in England*'. House of Commons Library. Available at: https://researchbriefings.files.parliament.uk/documents/CBP-9710/CBP-9710.pdf (accessed 15 February 2024).

McCullouch, J. (2016) 'The Contribution of Teacher Confidence to 'Excellent' Mathematics Teaching'. *Philosophy of Mathematics Education*, 30, 1–12. https://education.exeter.ac.uk/research/centres/stem/publications/pmej/pome30/index.html

Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., Tawell, A., Leseman, P. and Broekhuisen, M. (2015) '*A Review of Research on the Effects of Early Childhood Education and Care (ECEC) Upon Child Development. CARE Project; Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC)*'. Available at: http://ecec-care.org/fileadmin/careproject/Publications/reports/new_version_CARE_WP4_D4_1_Review_on_the_effects_of_ECEC.pdf (accessed 8 October 2019).

Melhuish, E., Howard, S.J., Siraj, I., Neilsen-Hewett, C., Kingston, D., de Rosnay, M., Duursma, E. and Luu, B. (2016) 'Fostering Effective Early Learning (FEEL) Through a Professional Development Programme for Early Childhood Educators to Improve Professional Practice and Child Outcomes in the Year Before Formal Schooling: Study Protocol for a Cluster Randomised Controlled Trial. *Trials*, 17: 1, 602. https://doi.org/10.1186/s13063-016-1742-1

Melhuish, E.C., Phan, M.B., Sylva, K., Sammons, P., Siraj-Blatchford, I. and Taggart, B. (2008) Effects of the Home Learning Environment and Preschool Center Experience upon Literacy and Numeracy Development in Early Primary School'. *Journal of Social Issues*, 64: 1, 95–114. https://doi.org/10.1111/j.1540-4560.2008.00550.x

Meyers, D.C., Durlak, J.A. and Wandersman, A. (2012) 'The Quality Implementation Framework: A Synthesis of Critical Steps in the Implementation Process'. *American Journal of Community Psychology*, 50: 3–4, 462–480. https://doi.org/10.1007/s10464-012-9522-x

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H. and Earl, L. (2014) 'State of the Art – Teacher Effectiveness and Professional Learning'. *School Effectiveness and School Improvement*, 25: 2, 231–256. https://doi.org/10.1080/09243453.2014.885451

Muijs, D., Reynolds, D., Sammons, P., Kyriakides, L., Creemers, B.P.M. and Teddlie, C. (2018) Assessing Individual Lessons Using a Generic Teacher Observation Instrument: How Useful is the International System for Teacher Observation and Feedback (ISTOF)? *ZDM Mathematics Education*, 50: 3, 395–406. https://doi.org/10.1007/s11858-018-0921-9

Mulligan, J. and Mitchelmore, M. (2009) 'Awareness of Pattern and Structure in Early Mathematical Development'. *Mathematics Education Research Journal*, 21: 2, 33–49. https://doi.org/10.1007/BF03217544

National Council of Teachers of Mathematics. (2014) '*Principles to Actions: Ensuring Mathematics Success for All*'. Reston, VA: National Council of Teachers of Mathematics.

Nguyen, T., Watts, T.W., Duncan, G.J., Clements, D.H., Sarama, J.S., Wolfe, C. and Spitler, M.E. (2016) 'Which Preschool Mathematics Competencies are Most Predictive of Fifth Grade Achievement?' *Early Childhood Research Quarterly*, 36, 550–560. https://doi.org/10.1016/j.ecresq.2016.02.003

Nunes, T., Barros, R., Evangelou, M., Strand, S., Mathers, S. and Sanders-Ellis, D. (2018) '*1stClass@Number: Evaluation Report and Executive Summary*'. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/public/files/1stClass@Number_evaluation_report.pdf (accessed 5 December 2019).

Office of National Statistics (ONS). 2023 School Workforce in England. Available at: **https://explore-education-statistics.service.gov.uk/find-statistics/school-workforce-in-england/2024** (accessed 12 February 2024)

Office for Standards in Education, Children's Services and Skills (Ofsted). (2019) '*The Education Inspection Framework No 190015*'. London: Ofsted. Available at:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/801429/Education_inspection_framework.pdf (accessed 20 January 2021).

Office for Standards in Education, Children's Services and Skills (Ofsted). (2021) '*Education Recovery in Schools: Autumn 2021'*. London: Ofsted. Available at: www.gov.uk/government/publications/education-recovery-in-schools-autumn-2021/education-recovery-in-schools-autumn-2021 (accessed 15 February 2024).

Organisation for Economic Co-operation and Development (OECD). (2013) '*OECD Skills Outlook 2013: First Results from the Survey of Adult Skills'*. Paris: Organisation for Economic Co-operation and Development. Available at: www.oecd-ilibrary.org/education/oecd-skills-outlook-2013_9789264204256-en (accessed 21 January 2021).

Pianta, R.C. (2012) '*Handbook of Early Childhood Education'*. London: Guildford Press.

Quigley, A., Muijs, D. and Stringer, E. (2021) *Metacognition and Self-Regulated Learning. Guidance Report*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/metacognition (accessed 12 February 2024).

Ritchie, S.J. and Bates, T.C. (2013) Enduring Links from Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status'. *Psychological Science*, 24: 7, 1301–1308. https://doi.org/10.1177/0956797612466268

Robinson-Smith, L., Fairhurst, C., Stone, G., Bell, K., Elliott, L., Gascoine, L., Hallett, S., Hewitt, C., Hugill, J., Torgerson, C., Torgerson, D., Menzies, V. and Ainsworth, H. (2018) '*Maths Champions: Evaluation Report and Executive Summary*'. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Maths_champions_evaluation_report.pdf (accessed 8 October 2019).

Rubin, D.B. (1987) '*Multiple Imputation for Nonresponse in Surveys'*. New York: John Wiley & Sons. https://doi.org/10.1002/9780470316696.fmatter

Saghaei, M. and Saghaei, S. (2011) 'Implementation of an Open-Source Customizable Minimization Program for Allocation of Patients to Parallel Groups in Clinical Trials'. *Journal of Biomedical Science and Engineering*, 4: 11, 734–739. https://doi.org/10.4236/jbise.2011.411090

Schaeffer, M.W., Rozek, C.S., Maloney, E.A., Berkowitz, T., Levine, S.C. and Beilock, S.L. (2021) 'Elementary School Teachers' Math Anxiety and Students' Math Learning: A Large-Scale Replication'. *Developmental Science*, 24: 4, p.e13080. https://doi.org/10.1111/desc.13080

Scharton, S. (2004) 'I Did It My Way: Providing Opportunities for Students to Create, Explain, and Analyze Computation Procedures'. *Teaching Children Mathematics*, 10: 5, 278–283. https://doi.org/10.5951/TCM.10.5.0278

Senn, S. (1994) Testing for Baseline Balance in Clinical Trials'. *Statistics in Medicine*, 13: 17, 1715–1726. https://doi.org/10.1002/sim.4780131703

Siegler, R.S. (2009) 'Improving the Numerical Understanding of Children From Low-Income Families'. *Child Development Perspectives*, 3: 2, 118–124. https://doi.org/10.1111/j.1750-8606.2009.00090.x

Siraj, I., Cheeseman, S. and Kingston, D. (2017) *'Quality Interactions Study (QIS) Final Report: Working With DoE Preschools to Strengthen the Quality of Intentional and Relational Pedagogy'*. Sydney: New South Wales Department of Education.

Siraj, I., Melhuish, E., Howard, S., Neilsen-Hewett, C., Kingston, D., de Rosnay, M., Duursma, E., Feng, X. and Luu, B. (2018a) '*Fostering Effective Early Learning (FEEL) Study: Final Report'*. Sydney: New South Wales Department of Education. Available at: www.earlychildhoodworkforce.org/Fostering%20Effective%20Early%20Learning%20Study (accessed 8 October 2019).

Siraj, I., Neilsen-Hewett, S., Melhuish, E., Kingston, D. and Luu, B. (2018b) '*Researching Environments for Early Learning (REEL) Study*'. Wollongong NSW: University of Wollongong.

Siraj-Blatchford, I., Sylva, K., Muttock, S., Gilden, R. and Bell, D. (eds.). (2002) '*Researching Effective Pedagogy in the Early Years*. *Research Report*'. Nottingham: Department for Education and Skills.

Siraj-Blatchford, Sylva, K., Taggart, B., Sammons, P., Melhuish, E. and Elliot, K. (2003) '*The Effective Provision of Pre-School Education (EPPE) Project : The Continuing Effects of Pre-school education at Age 7 Years*'. London: Institute of Education. Available at: https://discovery.ucl.ac.uk/id/eprint/10005290/ (accessed 19 November 2019).

Skibbe, L.E., Montroy, J.J., Bowles, R.P. and Morrison, F.J. (2019) 'Self-Regulation and the Development of Literacy and Language Achievement from Preschool Through Second Grade'. *Early Childhood Research Quarterly*, 46, 240–251. https://doi.org/10.1016/j.ecresq.2018.02.005

Stokes, L., Hudson-Sharp, N., Dorsett, R., Rolfe, H., Anders, J., George, A., Buzzeo, J. and Munro-Lott, N. (2018) '*Mathematical Reasoning: Evaluation Report and Executive Summary*'. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Mathematical_Reasoning.pdf (accessed 5 December 2019).

Torgerson, D.J. and Torgerson, C.J. (2008) '*Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction*'. Hampshire and New York: Palgrave Macmillan.

Vygotsky, L.S. (1978) '*Mind in Society: The Development of Higher Psychological Processes*'. Cambridge, MA: Harvard University Press.

Walker, J.M.T., Shenker, S.S. and Hoover-Dempsey, K.V. (2010) 'Why Do Parents Become Involved in Their Children's Education? Implications for School Counselors'. *Professional School Counseling*, 14: 1, 27–41. https://doi.org/10.1177/2156759X1001400104

Whitebread, D. (2012) '*Developmental Psychology and Early Childhood Education*'. London: SAGE Publications.

Wylie, C. and Hodgen, E. (2011) '*Forming Adulthood: Past, Present and Future in the Experiences and Views of the Competent Learners @ 20*'. Wellington, New Zealand: Ministry of Education. Available at: https://thehub.sia.govt.nz/assets/documents/40520_Forming_Adulthood_0.pdf (accessed 12 February 2024).

Appendix A: Changes since the previous evaluation

Appendix A Table 1: Comparison of the FEEL trial and TEEMUP trial design as in the protocol

	Feature	FEEL	TEEMUP
Intervention	Content	'Leadership for Learning PD' was designed to cover the foundational principles of child learning and development, including: self-regulation; language and communication; conceptual development in maths; and science and critical thinking. The PD featured a cascading model of delivery to prepare participants to take up a leadership role within their workplaces and share their new knowledge with colleagues and families' (Siraj *et al.,* 2018a)	Focus on developing children's mathematical understanding including problem-solving, thinking and argumentation, as well as support for children's behaviour for learning Additionally, aspects such as self-regulation, understanding disadvantage, support for cognitively challenging interactions and intentional and relational pedagogy also feature Partnership working with families
	Recipients	90 educators across 38 settings	Minimum of two teachers across 53 schools. Including at least 1 x Reception teacher and 1 x Year 1 teacher (no more than three in total), plus 1 x SMT can attend initial two-day training
	Delivery model	2 x full days and 5 x half-days of face-to-face workshops Online support and learning Cascaded to others within workplace	2 x full days, 8 x half-days of face-to-face workshops Specialists needs-based coaching/mentoring in schools (minimum three school visits). Website with PD resources and additional materials Delivered to Reception and Year 1 teachers No cascade, collaborative working between Reception and Year 1 encouraged
	Duration	7–9 months	16 months
	Location	Australia	England
Trial design	Type	Cluster RCT	Cluster RCT
	Point of randomisation	Randomised once centre environmental quality ratings were completed but before pupil recruitment and completion of pupil baseline assessments	After child recruitment and baseline data collection are completed
	Control condition	Business as usual with wait-list design (control centres received the intervention at the end of the trial)	Business as usual plus £750 with TEEMUP PD available to purchase at the end of the trial at a discounted rate, if found to be effective
Sample	Settings	Long-day care/or preschool Early childhood education and care centres *Main analyses: Total n*=90 (*n*=38 intervention group; *n*=40 control group main analyses *Alternative analyses:* Total *n*=95 (n=40 intervention; *n*=55 control)	State funded primary schools *Total n*=106 (*n*=53 intervention group; *n*=53 control group)
	Pupils	Children three to five years old *n*=1,346	Children four to six years old Total *n*=3180 Cohort 1 Reception children four to five years old in academic year 2021–2022 (pre-test) without significant SEND or EAL, followed until the end of

		Inclusion: Parent/carer opt-in consent	Year 1 2022–2023 when they are five to six years old (post-test) *n*=1,590 Cohort 2 Reception children four to five years old post-tested only at the end of the academic year 2022–2023 without significant SEND or EAL *n*=1,590 Inclusion: Completion of baseline assessment Exclusion: Parent/carer withdrawal, significant SEND or EAL
Outcomes and source	Primary	Environmental: Early Childhood Environment Rating Scale-Extended (ECERS-E) Sustained Shared Thinking and Emotional Well-being (SSTEW) Scale	Child math attainment: British Ability Scales 3 Early Number Concepts by GL Assessment at pre- and post-test with Cohort 1
	Secondary	Child language attainment: Differential Ability Scales (DAS II): Verbal Comprehension Early Years Toolbox Expressive Vocabulary assessment Child math attainment: DAS III Early Number Concepts DAS II Early Numeracy Assessment Preschool Early Numeracy Scale (PENS) Self-regulation and PSED: Child Self-regulation and Behaviour Questionnaire (CSBQ) Strengths & Difficulties Questionnaire	Child math attainment: British Ability Scales 3 Early Number Concepts by GL Assessment collected by trained, blinded research assistants at post-test – Cohort 2 only Number EYSFP ELG and Numerical Patterns EYFSP ELG. Combined. Cohort 1 and 2 Self-regulation and PSED: Child Self-regulation and Behaviour Questionnaire (CSBQ) Self-Regulation EYFSP Early Learning Goal (ELG) Child general attainment: All 17 EYFSP ELGs average total point score Good Level of Development met (if available). Cohort 1 or 2 Teacher confidence: maths: Adapted 'Early Math Beliefs and Confidence Survey' (Chen *et al.* 2014)
Statistical analysis		Environmental quality (as measured by ECERS-E and SSTEW) for settings in the intervention group and settings in the control group compared using regression, controlling for geography, service type, National Quality Standard (NQS) rating, area-level socio-economic status, baseline environment quality ratings	Maths attainment (as measured by BAS3 ENC) for children in the intervention group and those in the control group will be compared using a mixed effects linear regression model at the child-level. Group allocation, baseline BAS3 ENC score, and the minimisation factors (geographical location of school, FSM, and EAL) will be included as fixed effects in the model, and school as a random effect. Analysis will be on an ITT basis

Appendix B: EEF cost rating

Appendix B Table 1: Cost rating

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix C: Security classification of trial findings

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	*Randomised design*	*<= 0.2*	*0-10%*					
4	*Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)*	*0.21 - 0.29*	*11-20%*					*4*
3	*Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)*	*0.30 - 0.39*	*21-30%*			*No adjustment for threats to internal validity*		
2	*Design for comparison that considers selection only on some relevant confounders*	*0.40 - 0.49*	*31-40%*					
1	*Design for comparison that does not consider selection on any relevant confounders*	*0.50 - 0.59*	*41-50%*					
0	*No comparator*	*>=0.6*	*>50%*					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	*Low*	*Adequate allocation sequence with concealed assignment*
Threat 2: Concurrent Interventions	*Low*	*The primary analysis model was rerun with the additional covariate of whether the school was partaking in the NCETM programme. Fifty-three schools (57.0%; Intervention, 25/47, 53.2%; Control, 28/49, 60.9%) reported taking part in this programme during the trial period. The sensitivity analysis showed that he estimated Hedges' g effect size was -0.07 (95% CI -0.05 to 0.19), which relates to 1 additional month's progress*

		in the intervention group. Since the effect size in this analysis is unchanged from the primary results, we conclude that any participation in the NCETM did not impact the TEEMUP results
Threat 3: Experimental effects	*Moderate*	*There is evidence of differential changes 60% of control schools did similar CPD. Sensitivity analyses accounting for these behaviours found effects similar to the main analysis.* *Usual practice data was collected and analysed. The report discusses the risk of contamination and concludes it did not bias the main findings.*
Threat 4: Implementation fidelity	*Moderate*	*About half of intervention schools were classified as "good compliers" (46.8% for Cohort 1, 57.5% for Cohort 2), with the rest being "minimal" or "non-compliant."* *Most schools engaged with the core components, but not all received the full "dose" of the intervention.* *Online versus face-to-face training taken up variably. Online described as 'light touch'. There was variation in attendance at training, mentor visits, and use of resources.* *Fidelity and compliance were well defined and measured. However, actual compliance was moderate: Only about half of schools met "good" compliance, and a notable proportion withdrew or were non-compliant.* *There were moderate deviations from the planned delivery, and not all participants received the full intervention as intended.*
Threat 5: Missing Data	*Low*	• *Pupil-level attrition for the primary outcome (BAS3 ENC maths attainment) was 15.1% (236 out of 1,566 pupils).* • *Intervention group: 14.9% attrition* • *Control group: 15.3% attrition* • *To investigate the impact of missing data, the primary analysis was repeated using MI by chained equations. The adjusted mean difference in BAS3 ENC score following MI was 0.50 (95% CI -0.74 to 1.73, p = 0.43) and the Hedges' g effect size was 0.05 (95% CI -0.07 to 0.16), which is similar to the ITT estimate. The multiple imputation analysis confirmed the robustness of the main results. Missing data did not materially affect the estimated impact of the intervention.*
Threat 6: Measurement of Outcomes	*Low*	*No baseline BAS3 ENC for Cohort 2, although not entirely necessary.*
Threat 7: Selective reporting	*Low*	*Transparent reporting.*

- **Initial padlock score [4]:** MDES of 0.17; attrition of 15.1%.
- **No adjustment made:** Up to two threats are classified as Moderate Risk and the direction of the likely biases is unknown or operates in opposite directions.
- **Final padlock score:** 4 Padlocks.

Appendix D: Effect size estimation

Appendix D Table 1: Effect size estimation, Cohort 1

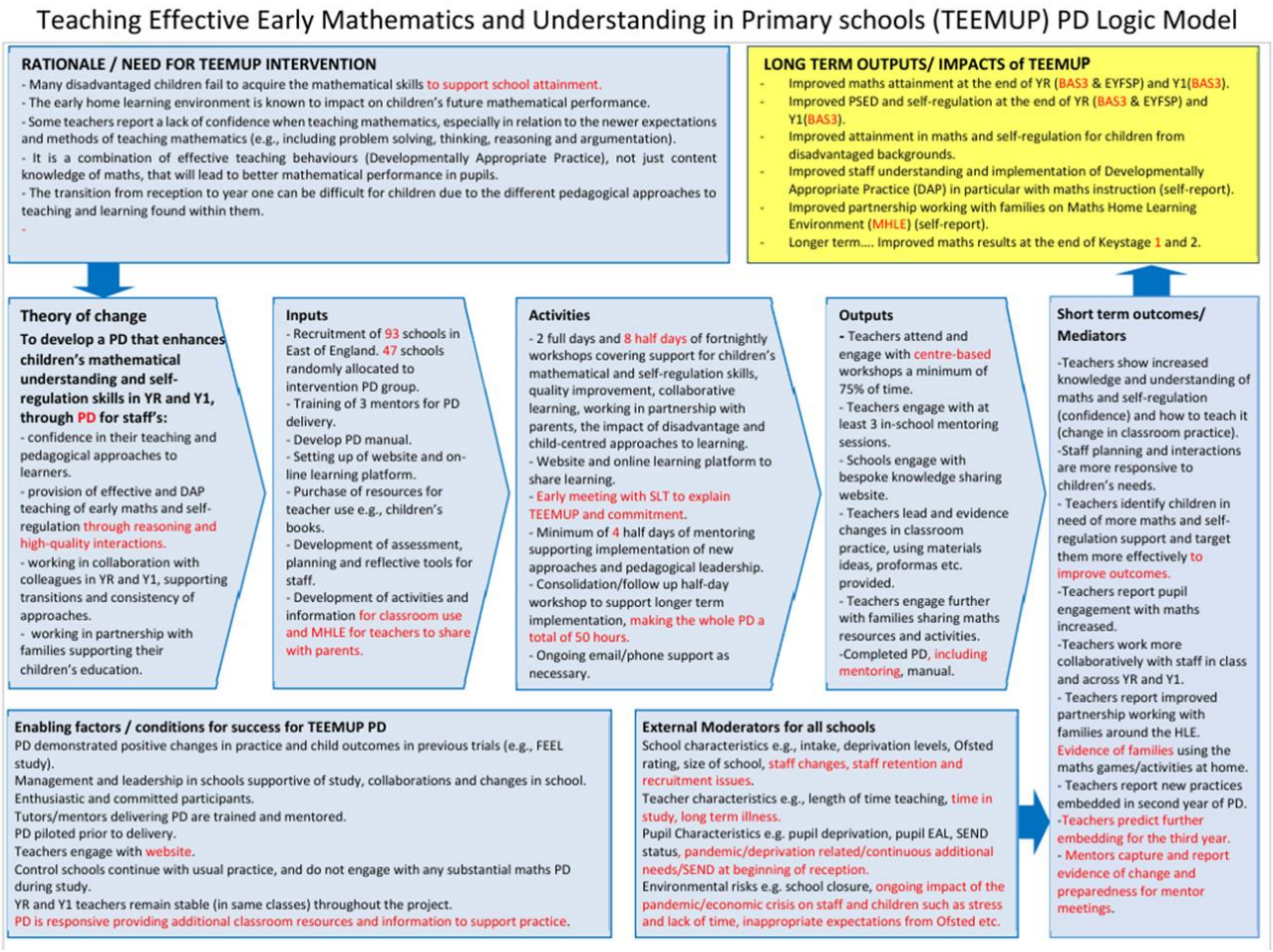
			Intervention group		Control group		
Outcome	Unadjusted differences in means	Adjusted differences in means	*n* (missing)	Variance of outcome	*n* (missing)	Variance of outcome	Pooled variance
BAS3 ENC	0.21 (-0.96, 1.38)	0.73 (-0.55, 2.02)	679 (101)	123.4	682 (104)	119.4	121.26
CSBQ Follow-up 1 (end of Reception)							
CSBQ Sociability	0.24 (0.15, 0.33)	0.16 (0.05, 0.26)	593 (382)	0.48	563 (416)	0.65	0.59
CSBQ Externalising	-0.22 (-0.31, -0.13)	-0.23 (-0.34, -0.12)	593 (382)	0.47	563 (416)	0.76	0.63
CSBQ Internalising	-0.28 (-0.35, -0.21)	-0.28 (-0.40, -0.16)	592 (383)	0.22	563 (416)	0.57	0.41
CSBQ Prosocial	0.26 (0.17, 0.35)	0.11 (-0.01, 0.23)	592 (383)	0.57	563 (416)	0.77	0.70
CSBQ Behavioural	0.31 (0.21, 0.41)	0.21 (0.08, 0.33)	593 (382)	0.62	563 (416)	0.88	0.77
CSBQ Cognitive	0.33 (0.23, 0.43)	0.21 (0.07, 0.35)	593 (382)	0.74	563 (416)	0.93	0.86
CSBQ Emotional	0.27 (0.18, 0.36)	0.17 (0.04, 0.29)	593 (382)	0.54	563 (416)	0.78	0.68
CSBQ Self-regulation	0.30 (0.21, 0.39)	0.19 (0.07, 0.30)	593 (382)	0.44	563 (416)	0.67	0.58
CSBQ Follow-up 2 (end of Year 1)							
CSBQ Sociability	0.06 (-0.02, 0.14)	-0.06 (-0.17, 0.05)	709 (266)	0.63	659 (320)	0.50	0.57
CSBQ Externalising	-0.14 (-0.23, -0.05)	-0.12 (-0.24, -0.01)	709 (266)	0.65	659 (320)	0.66	0.66
CSBQ Internalising	-0.15 (-0.23, -0.07)	-0.11 (-0.24, 0.01)	709 (266)	0.48	659 (320)	0.60	0.54
CSBQ Prosocial	0.03 (-0.06, 0.12)	-0.12 (-0.25, -0.00)	709 (266)	0.77	659 (320)	0.66	0.72
CSBQ Behavioural	0.05 (-0.05, 0.15)	-0.07 (-0.20, 0.07)	709 (266)	0.98	659 (320)	0.86	0.93
CSBQ Cognitive	0.05 (-0.06, 0.16)	-0.05 (-0.20, 0.10)	709 (266)	1.10	659 (320)	0.98	1.04
CSBQ Emotional	0.14 (0.05, 0.23)	0.04 (-0.09, 0.17)	709 (266)	0.76	659 (320)	0.64	0.71
CSBQ Self-regulation	0.08 (-0.01, 0.17)	-0.03 (-0.15, 0.09)	709 (266)	0.72	659 (320)	0.59	0.66
EYFSP ELG average point score	0.00 (-0.02, 0.02)	0.00 (-0.03, 0.03)	776 (4)	0.05	781 (5)	0.05	0.05

Appendix D Table 2: Effect size estimation, Cohort 2

			Intervention group		Control group		
Outcome	Unadjusted differences in means	Adjusted differences in means	*n* (missing)	Variance of outcome	*n* (missing)	Variance of outcome	Pooled variance
BAS3 ENC	0.08 (-1.19, 1.03)	0.08 (-1.28, 1.45)	525 (0)	79.6	476 (0)	80.1	79.7
CSBQ Sociability	0.40 (0.29, 0.51)	0.24 (0.11, 0.37)	430 (54)	0.54	436 (49)	0.73	0.67
CSBQ Externalising	-0.22 (-0.32, -0.12)	-0.08 (-0.22, 0.06)	430 (54)	0.47	436 (49)	0.73	0.61
CSBQ Internalising	-0.32 (-0.41, -0.23)	-0.23 (-0.35, -0.10)	430 (54)	0.32	436 (49)	0.52	0.45
CSBQ Prosocial	0.36 (0.25, 0.47)	0.16 (0.01, 0.31)	430 (54)	0.64	436 (49)	0.79	0.75
CSBQ Behavioural	0.39 (0.27, 0.51)	0.17 (0.02, 0.32)	430 (54)	0.75	436 (49)	0.93	0.88
CSBQ Cognitive	0.43 (0.31, 0.55)	0.22 (0.08, 0.35)	430 (54)	0.76	436 (49)	0.98	0.92
CSBQ Emotional	0.36 (0.25, 0.47)	0.15 (0.01, 0.29)	430 (54)	0.59	436 (49)	0.77	0.71
CSBQ Self-regulation	0.39 (0.29, 0.49)	0.17 (0.04, 0.29)	430 (54)	0.52	436 (49)	0.66	0.63
EYFSP ELG average point score	0.03 (0.00, 0.06)	0.03 (0.01, 0.05)	546 (3)	0.05	510 (1)	0.07	0.06

Appendix E: Revised logic model

Appendix E Figure 1: Revised logic model (with revisions displayed in red)



Appendix F: IPE surveys data tables

Appendix F Table 1: Reception and Year 1 teachers' usual practice at the start and end of the trial

	Start of trial (Autumn 2021)			End of trial (Summer 2023)		
Thinking about your usual practice	Intervention (*n*=76)	Control (*n*=73)	Total (*n*=149)	Intervention (*n*=57)	Control (*n*=55)	Total (*n*=112)
I use formative assessment to monitor and inform teaching						
Weekly	50 (65.8)	48 (65.8)	98 (65.8)	40 (70.2)	30 (54.5)	70 (62.5)
Monthly	20 (26.3)	19 (26.0)	39 (26.2)	12 (21.1)	22 (40.0)	34 (30.4)
Every 6 months	3 (3.9)	6 (8.2)	9 (6.0)	2 (3.5)	2 (3.6)	4 (3.6)
Yearly	3 (3.9)	0 (0.0)	3 (2.0)	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.8)	1 (0.9)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	3 (5.3)	0 (0.0)	3 (2.7)
I use tools/scales to reflect on my own teaching practice						
Weekly	15 (19.7)	13 (17.8)	28 (18.8)	11 (19.3)	10 (18.2)	21 (18.8)
Monthly	13 (17.1)	10 (13.7)	23 (15.4)	24 (42.1)	16 (29.1)	40 (35.8)
Every 6 months	15 (19.7)	16 (21.9)	31 (20.8)	8 (14.0)	6 (10.9)	14 (12.5)
Yearly	6 (7.9)	11 (15.1)	17 (11.4)	4 (7.0)	9 (16.4)	13 (11.6)
Never	27 (35.5)	23 (31.5)	50 (33.6)	7 (12.3)	14 (25.5)	21 (18.8)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	3 (5.3)	0 (0.0)	3 (7.1)
I think about how I can improve the transition of pupils into their new year group:						
Rarely	4 (5.3)	5 (6.8)	9 (6.0)	0 (0.0)	0 (0.0)	0 (0.0)
Sometimes	22 (28.9)	23 (31.5)	45 (30.2)	3 (5.3)	10 (18.2)	13 (11.6)
Often	35 (46.1)	36 (49.3)	71 (47.7)	31 (54.4)	27 (49.1)	58 (51.8)
Always	15 (19.7)	9 (12.3)	24 (16.1)	20 (35.1)	18 (32.7)	38 (33.9)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	3 (5.3)	0 (0.0)	3 (2.7)

Appendix F Table 2: Reception and Year 1 teachers' engagement with the home learning environment at the start and end of the trial, by allocated group

Thinking about your usual practice	Start of trial (Autumn Term 2021)			End of trial (Summer Term 2023)		
	Intervention (*n*=76)	Control (*n*=73)	Total (*n*=149)	Intervention (*n*=57)	Control (*n*=55)	Total (*n*=112)
I engage with pupils' families to improve the home learning environment:						
Weekly	25 (32.9)	31 (42.5)	56 (37.6)	21 (36.8)	19 (34.5)	40 (35.7)
Monthly	27 (35.5)	23 (31.5)	50 (33.6)	15 (26.3)	13 (23.6)	28 (25.0)
Every six months	16 (21.1)	13 (17.8)	29 (19.5)	9 (15.8)	15 (27.3)	24 (21.4)
Yearly	7 (9.2)	3 (4.1)	10 (6.7)	8 (14.0)	6 (10.9)	14 (12.5)
Never	1 (1.3)	3 (4.1)	4 (2.7)	1 (1.8)	2 (3.6)	3 (2.7)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	3 (5.3)	0 (0.0)	3 (2.7)

Appendix F Table 3: Reception and Year 1 teachers' responses to intervention delivery surveys relating to the quality of the TEEMUP training workshops in Summer Term 2022 and Summer Term 2023

Likert rating, *n* (%)	Quality	
	Summer Term 2022 *n*=39	Summer Term 2023 *n*=57
Venue-based training workshops		
1 (poor)	0 (0.0)	0 (0.0)
2	0 (0.0)	0 (0.0)
3	0 (0.0)	0 (0.0)
4	0 (0.0)	1 (1.8)
5	0 (0.0)	0 (0.0)
6	2 (5.1)	1 (1.8)
7	3 (7.7)	6 (10.5)
8	13 (33.3)	11 (19.3)
9	7 (17.9)	12 (21.1)
10 (exceptional)	11 (28.2)	13 (22.8)
Missing	3 (7.7)	13 (22.8)
Mean (SD)	8.6 (1.2)	8.6 (1.3)
Online-based training workshops		
1 (poor)	0 (0.0)	0 (0.0)
2	0 (0.0)	0 (0.0)
3	0 (0.0)	0 (0.0)
4	1 (2.6)	1 (1.8)
5	2 (5.1)	1 (1.8)
6	6 (15.4)	6 (10.5)
7	10 (25.6)	8 (14.0)
8	6 (15.4)	9 (15.8)
9	1 (2.6)	5 (8.8)
10 (exceptional)	3 (7.7)	4 (7.0)
Missing	10 (25.6)	23 (40.4)
Mean (SD)	7.1 (1.5)	7.6 (1.5)

Appendix F Table 4: Reception and Year 1 teachers' responses to intervention delivery surveys relating to the quality and relevance/usefulness of the TEEMUP mentor support in Summer Term 2022 and Summer Term 2023

Likert rating	Quality		Relevance/usefulness	
	Summer Term 2022 (*n*=39)	Summer Term 2023 (*n*=57)	Summer Term 2022 (*n*=39)	Summer Term 2023 (*n*=57)
	n (%)			
Mentor support to date				
1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
2	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
3	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
4	0 (0.0)	1 (1.8)	0 (0.0)	1 (1.8)
5	0 (0.0)	0 (0.0)	3 (7.7)	1 (1.8)
6	1 (2.6)	3 (5.3)	1 (2.6)	3 (5.3)
7	3 (7.7)	1 (1.8)	2 (5.1)	2 (3.5)
8	8 (20.5)	7 (12.3)	11 (28.2)	10 (17.5)
9	14 (35.9)	14 (24.6)	14 (35.9)	14 (24.6)
10	11 (28.2)	24 (42.1)	8 (20.5)	19 (33.3)
Missing	2 (5.1)	7 (12.3)	0 (0.0)	7 (12.3)
Mean (SD)	8.8 (1.0)	9.0 (1.3)	8.4 (1.4)	8.7 (1.4)

Appendix F Table 5: Reception and Year 1 teachers' responses to intervention delivery surveys relating to the quality and relevance/usefulness of the TEEMUP website

Likert rating, *n* (%)	Quality		Relevance/usefulness	
	Summer Term 2022 (*n*=39)	Summer Term 2023 (*n*=57)	Summer Term 2022 (*n*=39)	Summer Term 2023 (*n*=57)
TEEMUP website resources				
1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
2	0 (0.0)	1 (1.8)	0 (0.0)	0 (0.0)
3	1 (2.6)	0 (0.0)	0 (0.0)	0 (0.0)
4	0 (0.0)	1 (1.8)	0 (0.0)	3 (5.3)
5	1 (2.6)	1 (1.8)	2 (5.1)	1 (1.8)
6	2 (5.1)	4 (7.0)	4 (10.3)	6 (10.5)
7	9 (23.1)	4 (7.0)	12 (30.8)	3 (5.3)
8	9 (23.1)	10 (17.5)	5 (12.8)	13 (22.8)
9	10 (25.6)	11 (19.3)	9 (23.1)	4 (7.0)
10	6 (15.4)	20 (35.1)	7 (17.9)	22 (38.6)
Missing	1 (2.6)	5 (8.8)	0 (0.0)	5 (8.8)
Mean (SD)	8.0 (1.5)	8.5 (1.8)	7.9 (1.5)	8.3 (1.8)

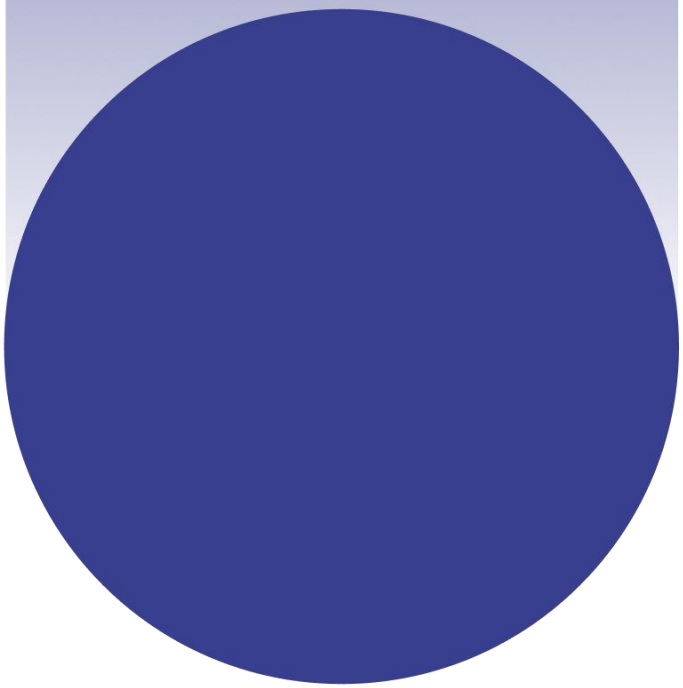
Appendix F Table 6: Reception and Year 1 teachers' responses to intervention delivery surveys relating to the quality and relevance/usefulness of the TEEMUP resources

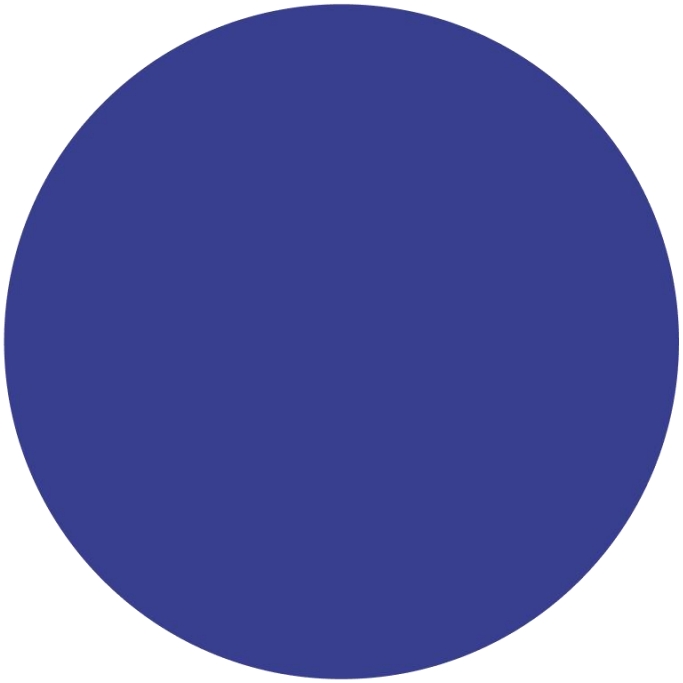
Likert rating, *n* (%)	Quality		Relevance/usefulness	
	Summer Term 2022 (*n*=39)	Summer Term 2023 (*n*=57)	Summer Term 2022 (*n*=39)	Summer Term 2023 (*n*=57)
Improving Maths Practice scale				
1–3*	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
4	0 (0.0)	1 (1.8)	0 (0.0)	3 (5.3)
5	0 (0.0)	3 (5.3)	2 (5.1)	4 (7.0)
6	2 (5.1)	3 (5.3)	3 (7.7)	3 (5.3)
7	5 (12.8)	5 (8.8)	6 (15.4)	4 (7.0)
8	5 (12.8)	14 (24.6)	9 (23.1)	16 (28.1)
9	19 (48.7)	13 (22.8)	12 (30.8)	12 (21.1)
10	7 (17.9)	13 (22.8)	7 (17.9)	10 (17.5)
Missing	1 (2.6)	5 (8.8)	0 (0.0)	5 (8.8)
Mean (SD)	8.6 (1.1)	8.3 (1.5)	8.2 (1.4)	8.0 (1.7)
Behaviour for Learning scale				
1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
2	0 (0.0)	0 (0.0)	1 (2.6)	1 (1.8)
3	0 (0.0)	1 (1.8)	0 (0.0)	1 (1.8)
4	0 (0.0)	0 (0.0)	1 (2.6)	2 (3.5)
5	2 (5.1)	1 (1.8)	3 (7.7)	3 (5.3)
6	2 (5.1)	9 (15.8)	3 (7.7)	5 (8.8)
7	7 (17.9)	5 (8.8)	6 (15.4)	6 (10.5)
8	3 (7.7)	16 (28.1)	8 (20.5)	14 (24.6)
9	17 (43.6)	8 (14.0)	10 (25.6)	10 (17.5)
10	7 (17.9)	11 (19.3)	7 (17.9)	9 (15.8)
Missing	1 (2.6)	6 (10.5)	0 (0.0)	6 (10.5)
Mean (SD)	8.4 (1.4)	8.0 (1.6)	7.8 (1.9)	7.7 (1.9)
Developmental progressions (learning trajectories)				
1–3*	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
4	0 (0.0)	0 (0.0)	1 (2.6)	1 (1.8)
5	2 (5.1)	2 (3.5)	2 (5.1)	4 (7.0)
6	4 (10.3)	1 (1.8)	2 (5.1)	3 (5.3)
7	4 (10.3)	9 (15.8)	6 (15.4)	6 (10.5)
8	11 (28.2)	8 (14.0)	14 (35.9)	8 (14.0)
9	7 (17.9)	11 (19.3)	5 (12.8)	11 (19.3)
10	10 (25.6)	20 (35.1)	9 (23.1)	18 (31.6)
Missing	1 (2.6)	6 (10.5)	0 (0.0)	6 (10.5)
Mean (SD)	8.2 (1.5)	8.7 (1.4)	8.1 (1.5)	8.4 (1.7)
Formative assessment materials				
1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
2	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
3	1 (2.6)	0 (0.0)	1 (2.6)	0 (0.0)
4	0 (0.0)	0 (0.0)	1 (2.6)	1 (1.8)
5	1 (2.6)	1 (1.8)	1 (2.6)	3 (5.3)
6	8 (20.5)	9 (15.8)	4 (10.3)	6 (10.5)
7	5 (12.8)	6 (10.5)	12 (30.8)	5 (8.8)
8	13 (33.3)	16 (28.1)	10 (25.6)	17 (29.8)
9	7 (17.9)	6 (10.5)	7 (17.9)	9 (15.8)
10	3 (7.7)	11 (19.3)	3 (7.7)	8 (14.0)
Missing	1 (2.6)	8 (14.0)	0 (0.0)	8 (14.0)
Mean (SD)	7.6 (1.5)	8.0 (1.5)	7.5 (1.5)	7.9 (1.5)

**Ratings condensed to reduce size of table.*

Further appendices:

Any further appendices are published as a separate document and can be found on the project page.





Education Endowment Foundation

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn