

**Effectiveness Trial of 1stClass@Number 1
Evaluation Protocol**

Evaluator (institution): National Foundation for
Educational Research
Principal investigator(s): Pippa Lord



Evaluation summary

Project title¹	Effectiveness trial of 1 st Class@Number 1
Developer <i>(Institution)</i>	Edge Hill University
Evaluator <i>(Institution)</i>	National Foundation for Educational Research
Principal investigator(s)	Pippa Lord
Protocol author(s)	Pippa Lord, Aarti Sahasranaman, Emma Moore, Chris Morton
Trial design	Two-arm randomised controlled trial with within school random allocation at pupil-level
Trial type	Effectiveness
Pupil age range and Key stage	6 – 7 years old, KS1
Number of schools <i>(at design stage)</i>	235
Number of pupils <i>(at design stage)</i>	1,880
Primary outcome measure and source	Quantitative Reasoning Test; test administered by teachers at baseline and by NFER test administrators at endpoint and data shared with NFER
Secondary outcome measure and source	NA

Protocol version history

Version	Date	Reason for revision
1.0 [original]	17/08/2023	N/A

Contents

Evaluation summary.....	1
Protocol version history.....	2
Study rationale and background.....	5
Intervention	6
Logic Model	11
Impact evaluation design	14
Implementation and process evaluation (IPE) design.....	25
Research questions	25
Cost evaluation design.....	36
Ethics and registration.....	37
Data protection.....	38
Personnel.....	39
Risks.....	40
Timeline	43
References	47
Appendix A: Changes since the previous EEF evaluation.....	48

Tables

Table 1: Trial design	14
Table 2: Sample size calculations	20
Table 3: Summary of the six compliance analyses to be performed	24
Table 4: IPE methods overview	35
Table 5: Timeline	43

Figures

Figure 1 Logic Model.....	13
---------------------------	----

Appendix tables

Appendix A: Changes since the previous EEF evaluation

Study rationale and background

1stClass@Number 1 is one of a suite of mathematical interventions developed by Every Child Counts², run on a not-for-profit basis by Edge Hill University (EHU) and set up with support from the Department for Education (DfE) in 2018. It is a light touch intervention for children who need further support at the level of the Year 1 curriculum and is delivered by trained Teaching Assistants (TAs) to small groups of four pupils. The 1stClass@Number interventions have been widely used by schools and so far have supported over 55,000 pupils in Years 1 to 11 across 4,000 schools³.

Every Child Counts (ECC) run their own internal evaluation of the 1stClass@Number interventions⁴ wherein schools are encouraged to submit the pre- and post-test outcome scores of pupils to ECC's online system, which, in turn, provides a detailed analysis of the scores. ECC recommend that schools delivering the 1stClass@Number interventions administer the standardised Sandwell Early Numeracy Test - Revised (SENT-R) at baseline and endpoint. Although not a rigorous Randomised Controlled Trial (RCT), this internal evaluation showed that pupils made, on average, a number age gain of 13 months in only 4 months. 93% of pupils who received the intervention showed more confidence and interest in learning mathematics in class. Given the positive impact demonstrated by ECC's internal evaluation and the widespread use of the 1stClass@Number interventions in schools, the Education Endowment Foundation (EEF) commissioned the evaluation of 1stClass@Number 1 through an effectiveness trial (Nunes *et al.*, 2018). This RCT with 130 schools found that pupils who received 1stClass@Number 1 made, on average, two additional months' progress in maths compared to pupils who did not receive the intervention. Although this finding had a high security rating, it was not statistically significant, likely due to inadequate sample size. There was no evidence to suggest that this progress in maths translated to improved KS1 maths outcomes. Among pupils eligible for Free School Meals (FSM), those who received the intervention did not make any additional progress in maths compared to those who did not receive the intervention. However, there was indicative evidence to suggest that 1stClass@Number 1 was not as effective for FSM pupils as for pupils not eligible for FSM. As part of the DfE's Accelerator Fund developed to close the disadvantage gap, the EEF subsequently commissioned a formative evaluation by the Behavioural Insights Team (BIT) to further explore this differential impact of 1stClass@Number 1 and to better understand the pupil selection mechanisms employed by schools⁵. The evaluators analysed pupils' pre- and post- test scores on the SENT-R collected by EHU and found no clear evidence that 1stClass@Number 1 was less beneficial to FSM pupils. The evaluation also found that schools selected a not unsubstantial proportion of pupils who were working above the recommended level in maths for the intervention (i.e. who did not appear to need additional support at the level of the Year 1 curriculum in terms of their age-related progress in maths, although we note that in a pandemic context, schools may also have been selecting using other criteria). The BIT evaluators, therefore, recommended that schools be provided with clear guidance on

² [About ECC - Every Child Counts \(edgehill.ac.uk\)](https://edgehill.ac.uk/about-ecc/)

³ [1stClass@Number - Every Child Counts \(edgehill.ac.uk\)](https://edgehill.ac.uk/1stclass@number/)

⁴ [1stClass@Number - Every Child Counts \(edgehill.ac.uk\)](https://edgehill.ac.uk/1stclass@number/)

⁵ The findings from this study were reported in an internal presentation to the EEF that was shared with potential evaluators at the time of submission of the expression of interest to deliver this evaluation.

selecting pupils to ensure the selection of pupils who were most likely to benefit from 1stClass@Number 1.

In light of evidence from the previous effectiveness trial and the findings of the formative evaluation, the EEF has commissioned this effectiveness trial to rigorously assess the impact of 1stClass@Number 1, with a particular focus on FSM-eligible pupils. Although our primary research question is to assess the impact of 1stClass@Number 1 on all pupils (to align with the EEF's remit across all trials), this trial is designed to be powered for the FSM subgroup, to enable us to assess the impact of the intervention on this subgroup. In response to the BIT's recommendation that schools be provided with clear guidance on selecting pupils, this trial will also implement a more objective selection process (detailed in the impact evaluation section of this protocol). Finally, one of the shortcomings of the earlier effectiveness trial was that it was underpowered, and we note that the findings were not statistically significant. The power calculations for that trial assumed a pre-test/post-test correlation of 0.7, whereas the correlation for nominated pupils in that trial was found to be more modest at 0.29. Assuming a pre-test/post-test correlation of 0.29, our power calculations precluded randomisation at the school-level as this would require a very large number of schools to be recruited (over 500), rendering the trial practically and financially infeasible. We have, therefore, designed this trial with pupil-level randomisation within schools. Our Implementation and Process Evaluation (IPE) will complement the impact evaluation and focus on areas such as the pupil selection guidance, training model and the pupil randomised design and the key mediators and moderators to understand the mechanisms by which the intervention will produce outcomes. Given the focus on FSM pupils in the impact evaluation, the IPE will include some lines of enquiry to explore any perceived differences for disadvantaged pupils (e.g. in terms of implementation, engagement, perceived outcomes), although we note that the intervention itself is for any pupil below expected standards (i.e. it is not an intervention specifically targeted at FSM pupils).

Intervention

The 1stClass@Number 1 intervention was developed by Every Child Counts (ECC), run on a not-for-profit basis by Edge Hill University and set up with support from the Department for Education (DfE). A detailed description of the intervention in the context of the TIDieR checklist is presented in the table below.

TIDieR Item	Description
Name	1 st Class@Number 1
Why (Theory/Rationale)	Small group maths tuition has been shown to be effective for improving pupils' attainment, particularly when targeted to pupils' specific needs. 1stClass@Number 1 was designed as a light-touch intervention to help pupils with moderate difficulties in maths catch-up with their peers. The intervention, delivered by Teaching Assistants (TAs) is based on connected learning experiences and the idea that children develop their mathematical understanding through: 1. Concrete experiences 2. Language 3. Pictures

	4. Symbols (e.g., those used for numbers, operations and equality) The programme covers five key topic areas: 1. All about number 2. Exploring place value 3. Addition and subtraction (1) 4. Addition and subtraction (2) 5. Towards multiplication and division These topic areas align to the topic areas in the National Curriculum and are designed to address the most common errors and misconceptions in primary maths learning.
Who (Recipients)	1stClass@Number 1 is targeted at Year 2 pupils (age 6 – 7 years) with moderate difficulties in mathematics to enable them to catch-up with their peers. Pupils are eligible for the intervention if they are achieving at around 12 months below age-related expectations at the end of Year 1. They may typically: • be able to count forwards in ones to 10 • have some knowledge of number facts and have some understanding of the composition of number within 10 • be able to perform simple addition and subtraction calculations using counting all and counting out approaches • be able to read and write numbers but lack secure understanding of their magnitude and quantity • struggle with mathematical vocabulary • lack confidence in mathematics and be reluctant to talk about their mathematical learning. Pupils nominated by schools to participate in the trial will be screened using the Sandwell Early Numeracy Test – Revised (SENT-R) and pupils scoring 40 or below will be eligible to participate in the intervention⁶. Pupils selected for 1stClass@Number 1 should not participate in any other maths interventions at the time 1stClass@Number 1 will be delivered. This trial is specifically focused on understanding the impact of 1stClass@Number 1 for disadvantaged pupils in receipt of Free School Meals (FSM). Schools are, therefore, encouraged to nominate FSM pupils who meet the criteria outlined above. Edge Hill University, the delivery partners, are recruiting schools to participate in this trial. To participate in the trial, schools are expected to: • be an Infant or Primary school with children in Year 2, as at 1st September 2023

⁶ Refer to the 'Pupil selection for trial' section on p.17 for details on how the SENT-R scores will be used to select pupils for the trial.

	• NOT be taking part in the following EEF-funded programmes in the academic years 2023-24 - ARK Mathematics Mastery or Mathematical Reasoning • NOT have a TA currently working in the school who has already completed any 1stClass@Number 1 training • NOT have delivered 1stClass@Number 1 after 1st September 2019 • nominate a TA and a Link teacher who are able to access face-to-face training in one of the 16 training venues • pay the subsidised training cost of £200 (usual cost is £1,100).
What (Materials)	TAs receive a handbook to help them organise their time during the weeks in which they deliver the intervention, detailed lesson plans for each topic, and most of the required materials to deliver the sessions (the only materials not included are those already commonly available in primary classrooms). ‘Special Delivery’, which consist of games/activities that pupils can take home to do with their parent/carer or another family member, are also included as part of the intervention.
What (Procedures, activities and/or processes)	TAs complete six half-day sessions of professional development delivered by Every Child Counts (ECC) trainers accredited by Edge Hill University and taking place every 2-4 weeks. These training sessions run alongside the intervention, so that TAs are trained in the intervention topic by topic. Training sessions for the TAs focus on how the intervention works, how to deliver it to pupils struggling with maths and developing TAs’ own understanding of maths. Link Teachers (LTs), who will provide strategic leadership of the programme in schools and support TAs, complete two half-day training sessions which focus on managing and supporting the implementation of 1stClass@Number 1 and sharing updates and learnings from the programme with colleagues. <u>Training model:</u> TAs receive 6 half-days of face-to-face training and LTs receive 2 half-days of training. Training for TAs covers five topics as detailed below. Each training day is followed by a 2 – 3 week break to allow for intervention delivery, following which training is provided in the next topic. In addition to preparation for the next topic, each TA training session also includes a feedback and review component where TAs can discuss their experience of delivering the previous topic to pupils and share any lessons learned. Training Day 1 consists of 2 half-day sessions – the morning session is common to TAs and LTs and provides an introduction to 1stClass@Number 1; the afternoon session is for TAs only and focuses on preparation for topic 1 (All about numbers). Training Day 2 prepares TAs to deliver topic 2 (Place value). Training Day 3 prepares TAs to deliver topics 3 and 4 (Addition and subtraction 1 and 2) and Day 4 prepares TAs to deliver topic 5 (Multiplication and division). Day 5 consists of one half-day session common to TAs and LTs where trainers discuss how to share the experiences of participating in the intervention at staff meetings in schools and how TAs and LTs could consider offering 1stClass@Number 1 for future Year 2 cohorts.

	Attendance at training sessions is monitored by EHU. For the purpose of the trial, trainers endeavour to offer a catch-up session (either face-to-face or remotely) to TAs and/or LTs who miss a training session. If this is not possible, then TAs can contact trainers to address any queries and use the topic handbook to deliver the intervention sessions. All schools will also receive one essential visit by the trainer generally meant to take place before the midpoint of the intervention to offer support and help quality assure delivery. Training and support visits are provided by 1stClass@Number trainers, who are experienced former teachers with expertise in the delivery of maths professional development to school colleagues. Trainers are also available for TAs and/or LTs to contact with any queries by telephone or email throughout the intervention period. It is expected that each trainer will be responsible for a group of 10-15 TAs, who would come together for each training session. The total number of training groups based on the number of schools required is expected to be 16 (and this will be delivered by 12 trainers)
Who (Intervention providers/implementers)	1stClass@Number 1 is one of a suite of maths interventions delivered by Every Child Counts and Edge Hill University. The intervention will be delivered by TAs who attend training. Link teachers are expected to meet with TAs once a week to help them review and plan upcoming lessons and provide feedback. The 1stClass@Number 1 team provide advice and guidance on the selection of TAs to implement the intervention. TAs should: • have successful experience of supporting children's mathematics, • be able to engage fully in training sessions that include the programme's procedures and the mathematics curriculum and • be able to make independent decisions while planning for sessions and teaching children based on an understanding of their needs and with the support of a Link Teacher • attend ALL six half-day training sessions The Link Teacher should be a senior member of staff (ideally a member of the school's leadership team) who: • is able to provide strategic leadership for the programme • is suitably experienced to give support with teaching maths • has time to support the TA and liaise with the class teacher, senior managers, SENCo and other professionals as necessary and • can attend BOTH half-day training sessions with the TA
How (Mode of delivery), when and how much (duration and dosage of the intervention)	1stClass@ Number 1 is designed to be delivered to small groups of four children. This group size allows the TA to observe and listen while children are completing activities in order to make necessary adjustments and enables a variety of pairings to be set, ensuring that children hear a variety of mathematical language. The intervention consists of 30 half-hour lessons, normally delivered 3 times a week for 10 weeks, during normal school hours in addition to regular

	mathematics instruction. This means that pupils involved in the intervention spend approximately 15 additional hours on numeracy. (Note, the training days are interspersed between topics, meaning the full time period of the whole intervention (training plus delivery) is approximately 15 weeks.) Each of the five topics is delivered in six sessions typically over a 2-week period. The sessions all follow the same structure starting with an introductory activity that sets the scene for both the TA and pupils in terms of what will be covered in the session. This is typically followed by a counting activity, followed by a specific area of mathematical focus and vocabulary for each session. The intervention handbook provides advice and guidance about allowing TAs sufficient preparation time and how to timetable the sessions for pupils. If specific pupils within the intervention group are struggling, then there is the option for TAs to deliver additional sessions to these pupils to help them keep up with the group. TAs will be asked to track pupils' attendance at each session and share these data with the evaluators. During this trial, the intervention will be delivered by TAs in schools from November 2023 – May 2024. The first cohort of recruited schools will deliver the intervention from November 2023 to March 2024 and the second cohort of recruited schools will deliver the intervention from January to May 2024.
Where (Location of the intervention)	For the purpose of this trial, it was planned that schools would be recruited from 16 regions across England – the regions originally planned included Birmingham, North Manchester, South Manchester, Liverpool, Lancashire, Rotherham, Cambridgeshire, East Sussex, Northeast, Peterborough, Wiltshire/Berkshire, Derbyshire/Nottinghamshire, North Somerset/South Gloucestershire, Dorset, Isle of Wight and a second training group covering different areas in Liverpool/Lancashire. Four of these training regions – Peterborough, Wiltshire/Berkshire, Isle of Wight and North Somerset/South Gloucestershire – were ultimately cancelled because of low demand and numbers were made up in other regions where demand for the programme was higher. A total of 5 schools recruited across these four regions have withdrawn from the trial and are being offered an online 1CN1 training programme by ECC outside of the trial. In the two cohort design, training groups that filled earlier were assigned to start training and intervention delivery in November 2023 (cohort 1) and those that are likely fill later will start intervention delivery in January 2024 (cohort 2). The following regions are included in the first cohort - Birmingham, North Manchester, South Manchester, Liverpool, Lancashire, Rotherham, Cambridgeshire, East Sussex, Derbyshire/Nottinghamshire, West Sussex and the second training group in Liverpool/Lancashire. The second cohort that is current being filled includes the Northeast, North Manchester (second group), Birmingham (second group), Derby and Dorset. These training areas have been chosen based on the availability of ECC trainers and some regions are also Education Investment Areas, regions of focus for recruitment in this trial. Ideally 1stClass@Number 1 sessions are delivered in a separate, designated space in participating schools where the group can sit together around a table without additional distractions, and where the noise they make during the games/activities will not disturb others.

Tailoring (Adaption of the intervention)	1 st Class@Number 1 comes ready-made with detailed session guidance and extensive resources. The lessons focus on number and calculation, developing children's mathematical understanding, communication and reasoning skills. Stimulating, enjoyable games and activities engage the children and build their confidence. Each topic starts with a simple assessment that helps the TA to tailor sessions to the children's needs. All the children still complete the same activities, but the TA can adapt them to make them easier or more difficult by for example using bigger or smaller numbers depending on the child's needs.
Strategies to support implementation	• TAs attend six small group training sessions on the course content. The first 5 sessions are roughly matched to the pace of delivery. • TAs are supported by LTs who also quality assure delivery. • Trainers also provide additional support in the form of an essential school visit to quality assure delivery and telephone and email support in case of queries. • Parents/carers are provided with information about the programme and 'Special Delivery' activities that pupils can take home to do with their parent/carer. • All consumable resources can be downloaded from the ECC website and used to deliver 1stClass@Number 1 to selected pupils in future Year 2 cohorts.

Logic Model

The logic model for 1stClass@Number 1 is shown in **Figure 1**. It outlines the target population (Year 2 pupils struggling with maths) of the intervention and the activities, outputs, short-term and intermediate outcomes that lead to the outcome of improving maths attainment both immediately after the intervention and longer term. The intervention involves a training programme for TAs and LTs as well as the intervention itself, so we have included each group of participants who take part in the intervention activities as a separate row within the logic model. This categorisation allows us to represent the different moderators that are applicable to each participant group and could have an impact on the overall attainment outcome. We have included both short-term and intermediate outcomes, to denote the ordering effect in which outcomes occur. These have not all been widely explored in previous research, so are a key area of interest for this trial. The IPE will explore these short-term and intermediate outcomes in more detail alongside the moderators that may lead to differential outcomes for FSM pupils/all pupils.

The intervention is mapped to the school curriculum and is designed to address the most common maths difficulties that pupils experience at this stage of their education. A key assumption of the logic model, therefore, is that the right pupils are selected for the programme. The programme provides guidance to schools on pupil selection, including recommending using a screening test (SENT-R) to aid selection. In response to BIT's formative study, we have amended this guidance for the purposes of the trial such that the SENT-R screening test is mandatory, thereby making the pupil selection process more objective.

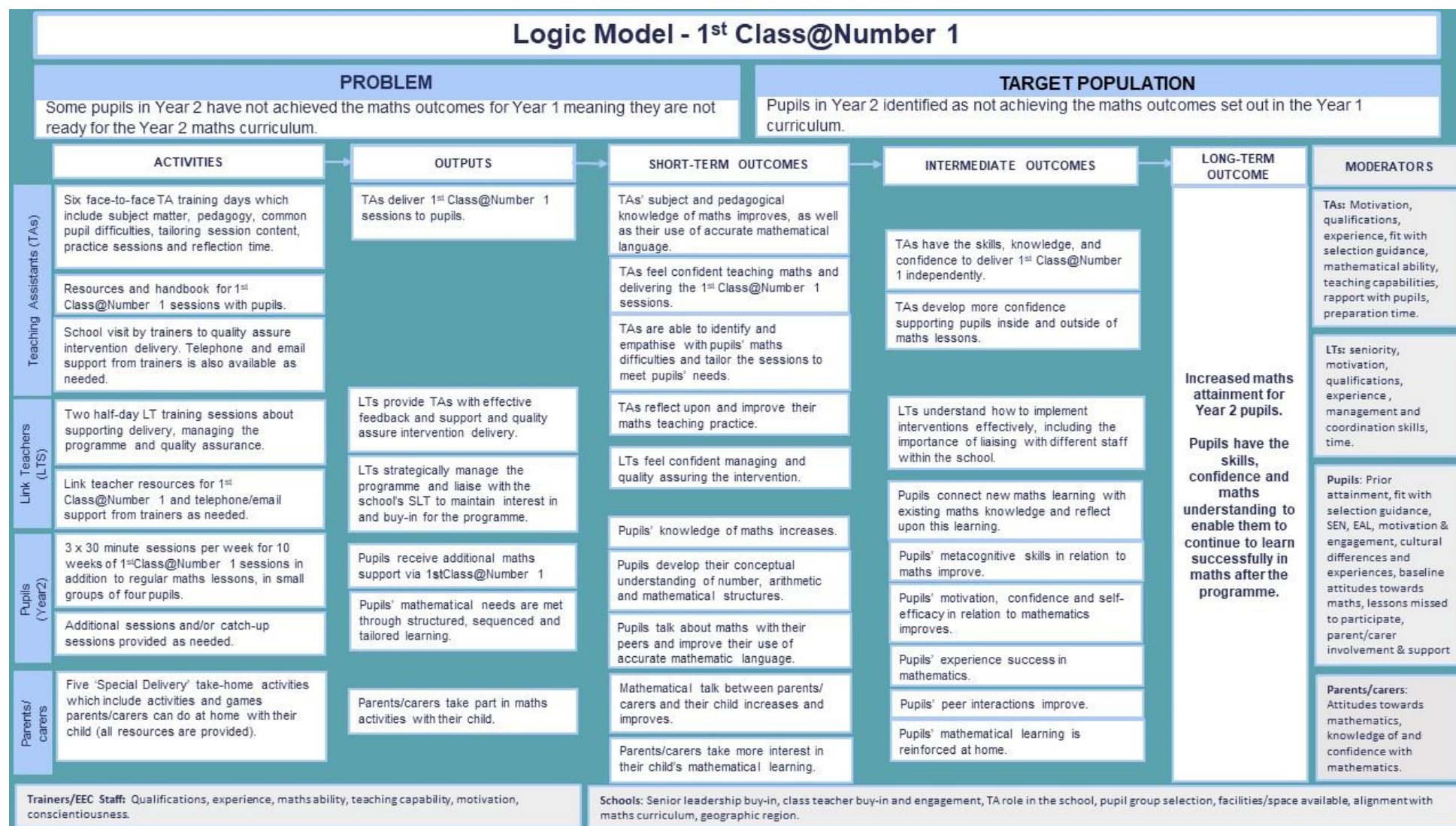
TA selection is also an important moderator, represented in the TIDieR checklist in terms of TA's prior experience in supporting maths and their ability to engage with the maths curriculum and subject knowledge in the training sessions themselves. It is expected that their subject and pedagogical knowledge of maths will improve during the intervention, and that they will be confident to deliver 1stClass@Number 1 and tailor sessions to their pupils (short-term outcomes). As set out in the logic model, it is then expected that this will lead to increased confidence to support pupils in maths more widely (intermediate outcomes).

The training TAs receive to deliver the intervention takes place alongside the intervention delivery. Therefore, the logic model makes the following assumptions 1) that the training delivery is consistent and high-quality, 2) that TAs will attend all of the training sessions and 3) that the time they have between the training session and delivering the relevant sessions for that topic is sufficient. Based on evidence from the previous trial of 1stClass@Number 1 (Nunes *et al.*, 2018) it appears that these assumptions are accurate, but we will check these three assumptions again in the fidelity, quality and compliance measures for this trial, as it involves a different group of participants.

In addition, there are also several assumptions relating to pupils' attitudes and behaviours that underlie the logic model outcomes. One of these is that increasing pupils' mathematical knowledge and conceptual understanding during the intervention (short-term outcomes) is sufficient to help them achieve success and feel more confident in maths (intermediate outcomes). Another assumption is that the combination of the small group setting, which gives pupils the opportunity to talk with each other about mathematics, and the Special Delivery activities, which are designed to encourage mathematical talk with their family, will lead to improved peer interactions more generally. We propose to explore how pupils feel about the intervention, their motivation and confidence in maths, and the difference they feel the intervention has made for them as part of the IPE.

There is currently very limited evidence as to how much of a role parents/carers play in the success of the programme and to what extent the Special Delivery activities are used at home. In addition, one of the short-term outcomes in the logic model is that mathematical talk increases between parents/carers and their child. This assumes that the pupil has someone at home that they can talk to about school and specifically maths, but this may not be the case for all pupils. It also assumes that the Special Delivery activities will encourage parents/carers to use mathematical talk with their child(ren). There is potential for increased parent/carer engagement to benefit pupils' mathematical learning as outlined in the logic model, but there are also potential moderators on any effects. We therefore plan to investigate perceptions of the role of parents/carers in this intervention as an exploratory part of the IPE when considering the context, barriers, and facilitators of this intervention.

Figure 1 Logic Model



Impact evaluation design

Research questions

The primary research question is:

- RQ 1. What is the difference in maths attainment measured by the Quantitative Reasoning Test of pupils receiving the 1stClass@Number 1 intervention in comparison to control pupils who do not receive the intervention?

The secondary research questions are:

- RQ 2.1. What is the difference in maths attainment measured by the Quantitative Reasoning Test of FSM-eligible pupils receiving the 1stClass@Number 1 intervention in comparison to FSM-eligible control pupils who do not receive the intervention?⁷
- RQ 2.2. Is the 1stClass@Number 1 intervention as effective for pupils with Special Education Needs (SEN) as for pupils without SEN as assessed by the Quantitative Reasoning Test?
- RQ 2.3. Is the 1stClass@Number 1 intervention as effective for pupils with lower prior attainment as for pupils with higher prior attainment as assessed by the Quantitative Reasoning Test?

Design

Table 1: Trial design

Trial design, including number of arms		Randomised controlled trial with two arms
Unit of randomisation		Within school randomisation at pupil-level
Stratification variables (if applicable)		FSM eligibility status
Primary outcome	Variable	Maths attainment
	Measure (instrument, scale, source)	Quantitative Reasoning Test (Terezinha Nunes, Peter Bryant, Deborah Evans & Rossana Barros (2015) Assessing Quantitative Reasoning in Young Children, Mathematical Thinking and Learning, 17:2-3, 178-196, DOI: 10.1080/10986065.2015.1016815)
Secondary outcome(s)	Variable(s)	Maths attainment as per primary outcome but with FSM pupils only
	Measure(s) (instrument, scale, source)	Quantitative Reasoning Test (Terezinha Nunes, Peter Bryant, Deborah Evans & Rossana Barros (2015) Assessing Quantitative Reasoning in Young Children, Mathematical Thinking and Learning, 17:2-3, 178-196, DOI: 10.1080/10986065.2015.1016815)

⁷ Although presented as a secondary research question, we are particularly interested in determining the impact of 1stClass@Number 1 on disadvantaged pupils, especially those in receipt of Free School Meals. Therefore, the design powers this trial to assess the impact of the intervention on FSM-eligible pupils.

Baseline for primary outcome	Variable	Maths attainment
	Measure (instrument, scale, source)	Quantitative Reasoning Test (Terezinha Nunes, Peter Bryant, Deborah Evans & Rossana Barros (2015) Assessing Quantitative Reasoning in Young Children, Mathematical Thinking and Learning, 17:2-3, 178-196, DOI: 10.1080/10986065.2015.1016815)
Baseline for secondary outcome	Variable	Maths attainment as per primary outcome but with FSM pupils only
	Measure (instrument, scale, source)	Quantitative Reasoning Test (Terezinha Nunes, Peter Bryant, Deborah Evans & Rossana Barros (2015) Assessing Quantitative Reasoning in Young Children, Mathematical Thinking and Learning, 17:2-3, 178-196, DOI: 10.1080/10986065.2015.1016815)

This is an effectiveness trial to assess the impact of the 1stClass@Number 1 intervention on maths attainment of Year 2 pupils, who will be the recipients of this intervention. We are particularly interested in the impact of this intervention on disadvantaged pupils, especially those in receipt of Free School Meals (FSM). Therefore, we have powered this trial to detect the impact on FSM-eligible pupils. The 1stClass@Number 1 intervention is offered to small groups of 4 pupils and randomisation will be within schools at the pupil-level. The decision to randomise at the pupil-level was made following practical considerations of running a very large school-randomised trial. Assuming a conservative pre-test/post-test correlation of 0.29 as determined in the previous effectiveness trial, our power calculations for a school-randomised trial powered for FSM-eligible pupils returned a sample size of 550 schools (assuming 50% of pupils selected to the trial are FSM-eligible). Through conversations with EHU, we established that these numbers far exceeded both their recruitment and delivery capacity if the trial were to be run for a single cohort during the academic year. We explored the possibility of running the trial in 2 cohorts over 2 academic years to address these recruitment and delivery challenges, but this option proved to be financially prohibitive for such a large number of schools. Having explored these options, we ultimately decided to pursue a pupil-randomised design within schools as working with the same number of schools in this design will give us more power when compared to a school-randomised design. Therefore, we will recruit 235 schools, randomising a total of 8 pupils from each school on a 1:1 basis into two arms: intervention and control. Given the focus on FSM pupils in this trial, randomisation will be stratified by pupils' FSM eligibility status to ensure a balance of FSM pupils in both arms. Pupils assigned to the intervention arm will receive the 10-week 1stClass@Number 1 intervention in addition to normal maths lessons. Pupils assigned to the control group will not receive the 1stClass@Number 1 intervention but will be able to access other support that may be provided to them as a matter of course. All schools will pay a subsidised training fee of £200 (the usual cost is £1,100) and will be eligible to receive an incentive payment of £500 upon completing all trial-related activities. The primary outcome for this trial will be maths attainment measured by the Quantitative Reasoning Test (QRT).

Recruitment of schools was scheduled to take place between February – June 2023. As recruitment progressed, it became evident that it would be difficult to reach the recruitment target of 235 schools by end of June 2023. Schools noted numerous challenges including the ongoing cost of living crisis, uncertainty over staffing and lack of budget that prevented them from signing up to participate in the trial. To ensure that the original target of 235 schools was recruited so that the trial could continue to be powered for FSM-eligible pupils, EEF, NFER and EHU agreed that a split cohort design would be the most feasible solution. In this revised

design, the trial will be run in the same academic year (2023/24) with the 235 schools split into two cohorts. Since recruitment by EHU takes place by training regions, training regions will be allocated to either cohort 1 or 2 depending on their likelihood of being filled earlier or later, respectively. Schools in cohort 1 will start receiving training and delivering the intervention in November 2023 and schools in cohort 2 will start receiving training and delivering the intervention in January 2024.

Participant selection

EHU, the intervention delivery partner, along with their local ECC providers will be responsible for recruiting 235 schools for this trial. They will do so through direct contact with schools by ECC trainers, media engagement, social media channels such as Twitter and other promotional events. EEF and NFER will support their recruitment efforts by promoting the trial through their newsletters and social media channels. Interested schools will complete an online Expression of Interest, which will help EHU ascertain schools' eligibility to participate in the trial. School eligibility criteria are set out below. In addition, schools should have high levels of FSM (at least 30%) and ideally 12 – 16 pupils suitable for nomination. Note, at EOI stage schools will be asked to complete whether they have at least 10 pupils whom they consider to be suitable for the trial (12 – 16 is the ideal, but at least 10 is in the EOI). Eligible schools will then be sent the school information sheet and Memorandum of Understanding (MoU). Schools will sign up to the trial by the headteacher signing on the MoU and providing the name of a key project contact who will act as the coordinator of the trial in the school. Once a school has signed up, EHU will share the school name and details of the headteacher and key project contact with NFER. NFER will get in touch with the key project contact to initiate data collection and other trial-related activities (see pupil eligibility section for further details). At this stage, where schools may not have 12 – 16 eligible pupils, they will be asked to nominate at least 8 eligible pupils for the trial, including an additional 2 pupils (i.e. a total of 10 pupils) who would benefit from the intervention. This would help us accommodate any change in circumstances of pupils so that 8 pupils are ultimately selected to participate in the trial.

School eligibility

In order to take part in this trial, schools must:

- be an Infant or Primary school with children in Year 2, as at 1st September 2023.
- NOT be taking part in the following EEF-funded programmes in the academic years 2023-24 - ARK Mathematics Mastery or Mathematical Reasoning
- NOT have a TA currently working in the school who has already completed any 1stClass@Number 1 training
- NOT have delivered 1stClass@Number 1 after 1st September 2019
- nominate a TA and a Link teacher who are able to access face-to-face training in one of the 16 training venues
- pay the subsidised training cost of £200 (usual cost is £1,100)

Note, as part of the Accelerator Fund, EHU also have a target for half of the recruited schools to be in Education Investment Areas (EIAs). In addition, because the trial requires at 50% of the pupils to be eligible for FSM, EHU will target schools with high levels of FSM during recruitment (e.g. with at least 30% FSM). EHU will monitor the %FSM in Year 1 at EOI and MoU stage. NFER will check the school level data received at MoU stage. During recruitment, should the average FSM across the recruited sample to date fall below 50%, NFER will ask EHU to re-target schools with even higher levels of FSM for the remainder of the recruitment.

Pupil eligibility

To identify pupils eligible to participate in this trial, we will follow a two-step process. In step 1, schools will be asked to nominate 12 – 16 pupils (see Participant Selection section above for further details on the number of pupils to be nominated by schools) that they feel are eligible to receive 1stClass@Number 1. This pupil nomination step will take place in the summer term of 2023 when pupils are in Year 1 and for some cohort 2 schools will take place in the autumn term of 2023 when pupils are in Year 2. In step 2 (September – November 2023), these nominated pupils will be screened using the Sandwell Early Numeracy Test - Revised (SENT-R) to identify the final 8 pupils who will be selected to participate in this trial. Pupil selection will take place in the autumn term of 2023 when pupils are in Year 2. It is our expectation that not all pupils who are nominated for screening will be eligible to receive 1stClass@Number 1 based on their SENT-R scores. This is why we are asking schools to nominate more pupils than the final 8 who will be selected to the trial.

Pupil Nomination for Screening

There are three broad eligibility criteria to determine which pupils can be nominated for screening:

1. The pupil should be achieving at around 12 months behind age-related expectations towards the end of Year 1.
2. The pupil should not be participating in another mathematics intervention when 1stClass@Number 1 is being delivered (autumn 2023/spring 2024).
3. This trial has a focus on the impact of 1stClass@Number 1 on disadvantaged pupils, especially those in receipt of Free School Meals. Schools are, therefore, encouraged to nominate FSM-eligible pupils who meet the above criteria.

Pupils who have fallen behind their peers and require additional support may typically:

- be able to count forwards in ones to 10
- have some knowledge of number facts and also have some understanding of the composition of number within 10
- be able to perform simple addition and subtraction calculations using counting all and counting out approaches
- be able to read and write numbers but lack secure understanding of their magnitude and quantity
- struggle with mathematical vocabulary
- lack confidence in mathematics and be reluctant to talk about their mathematical learning.

NFER and EHU have developed a guidance document for schools to enable them to choose the 12 – 16 most appropriate pupils for screening. The document also outlines how schools should proceed with pupil nomination in the event that they have fewer than 12 pupils to nominate. When schools are only able to nominate 8 pupils who meet the criteria above, our recommendation is that they nominate 2 additional pupils to accommodate any change in

circumstances of the 8 pupils thought to be eligible for screening. While these remaining 2 pupils may not fully adhere to the selection criteria above, they must be expected to benefit from the intervention. This guidance document will be shared by EHU with schools after they have signed the MoU.

NFER will contact schools after they have signed the MoU to share the parent letter and withdrawal form and the pupil data form. Schools will be asked to share the parent letter and withdrawal form with the parents/carers of all pupils that they intend to nominate for screening. At this stage, parents/carers will have the opportunity to withdraw their child from the evaluation and associated data processing. Once the withdrawals have been processed, schools will share pupil administrative data for the remaining pupils with NFER via the pupil data form. This will include pupil names, date of birth, unique pupil number (UPN), FSM eligibility and class name in Year 2 (if known). Schools will also be asked to share details of the TA and link teacher (names, role, email address, telephone number) they wish to nominate for training. Schools may not have decided on which staff members to nominate for training at this stage and will be given another opportunity in autumn 2023 to provide these data.

Pupil Selection for Trial

Once schools have nominated pupils for screening and shared their administrative data with NFER, the next step is to select the final 8 pupils for the trial. Nominated pupils from each school will undertake the SENT-R in September - November 2023 at the start of Year 2 to objectively confirm that they are achieving below age-related expectations. Schools will be asked to administer Form A of the SENT-R (the test kit provides two versions, Forms A and B to allow for pre- and post- measures). The SENT-R screening test will be administered 1:1 in schools by teachers and the test papers will be sent back to NFER for processing the scores. NFER will share a customised report with each school rank ordering pupils based on their raw SENT-R scores and recommending 8 pupils for selection to the trial. Schools will then be asked to confirm the final 8 participants to be selected to the trial.. In recommending pupils for selection to the trial, we will first prioritise the selection of FSM-eligible pupils with raw SENT-R scores of 49 or below. While pupils scoring 40 or below are considered to be around 12 months behind age-related expectations, in agreement with EHU we are also allowing the selection of pupils with a raw score between 41 – 49, where a raw score of 49 equates to a number age of 6:10 (this is part of usual practice that EHU experience). Once FSM-eligible pupils have been recommended, we will randomly choose non-FSM pupils with raw scores of 40 and below to reach the target of 8 pupils for recommendation to the trial. If this target is not achieved by choosing pupils with scores of 40 or below, we will then randomly choose non-FSM pupils with a raw score of 41 - 49 in order to reach the required number of 8 pupils. In the event that a school does not have 8 pupils who fall below the score of 49 on the SENT-R, we will then randomly choose pupils with scores above 49 to reach the target of 8 pupils (this is preferable to randomising fewer than 8 pupils in a school, further details below). Schools will be asked to briefly confirm the recommended shortlist, and also given the option of replacing recommended non-FSM pupils with other pupils who have not been recommended by us.

This selection mechanism will allow all recruited schools to continue participating in the trial even if they do not have 8 pupils who strictly meet the SENT-R eligibility criteria. The section on 'subgroup analyses' provides details on how we will manage the SENT-R range in our dataset, should some pupils fall above the raw score cut-off. An alternative approach we considered was that of unequal randomisation in the event that fewer than 8 pupils were eligible based on the strict SENT-R cut-off score of 40. However, having considered the

analytical issues due to more pupils always being allocated to the intervention arm (since the intervention must be delivered to groups of 4 pupils), we ultimately decided against this option. We may consider this option if fewer than 8 pupils in any school complete the SENT-R screening test.

In addition to the SENT-R, all nominated pupils will also complete a baseline QRT assessment within a 2 – 3 week assessment window in September 2023 (cohort 1) and October – November 2023 (cohort 2). The baseline QRT will be administered in schools by teachers to all nominated pupils as a group and the test papers will be sent back to NFER for marking. Once the screening and baseline assessment are complete, schools will confirm the list of 8 selected pupils with NFER. Completion of the screening and baseline assessment and confirming details of the 8 selected pupils are prerequisites for randomisation. Random assignment of the 8 pupils to the intervention or control arms in a 1:1 ratio will take place in October 2023 (cohort 1) and January 2024 (cohort 2). Details of which arm each pupil is randomly assigned to will be shared with schools to enable them to deliver the intervention to the appropriate pupils.

TA training and intervention delivery will take place from November 2023 to March 2024 for cohort 1 and from January 2024 to May 2024 for cohort 2. Upon completion of intervention delivery, the endpoint QRT assessments will be administered by NFER test administrators in schools in March 2024 (cohort 1) and May – June 2024 (cohort 2) and the test papers sent back to NFER for marking. After data collection for the trial has been completed, schools may have the option of delivering 1stClass@Number 1 to control pupils and/or other pupils who might be eligible for the intervention.

Outcome measures

Baseline measures

We will use the QRT score (Nunes *et al.*, 2015) in its entirety as the baseline measure. At baseline, QRT will be administered by teachers in schools in September 2023 (cohort 1) and October – November 2023 (cohort 2) and the test papers will be sent back to NFER for marking. Although the baseline QRT tests will be completed before randomisation (scheduled for October 2023 for cohort 1 and January 2024 for cohort 2), the tests will be marked after randomisation has been completed. The tests will be marked by trained NFER staff who are blinded to the arm to which pupils are randomly assigned. Baseline QRT scores will be included in the models as a covariate. This will be described in greater detail in the Statistical Analysis Plan (SAP).

Primary outcome

As in the previous effectiveness trial, the primary outcome for this trial is also maths attainment as measured by the QRT (Nunes *et al.*, 2015). At endpoint, QRT will be administered in schools (in March 2024 for cohort 1, and in May-June 2024 for cohort 2) by NFER test administrators who are blinded to the arm to which pupils are randomly assigned. The tests will be marked by trained NFER staff who are also similarly blinded.

The QRT comprises four scales that measure additive composition, the inverse relationship between addition and subtraction, additive reasoning and multiplicative reasoning and is a good match for the 1stClass@Number 1 intervention which aims to build pupils' confidence in number sense, place value, and calculation. The reliability and construct validity of the QRT

are demonstrated by the good alpha reliability levels of the four scales and test-retest correlation for the total score ($R=0.78$, $p<0.001$) (Nunes *et al.*, 2015). The four scales correlate significantly with each other, and each measure has a factor loading above 0.8. The QRT has been validated among young children aged 5 - 6 years and has been shown to be a strong predictor of KS1 attainment (correlation of 0.7 when the assessments were 14 months apart) (Nunes *et al.*, 2007).

Based on the logic model for 1stClass@Number 1, the intervention is expected to increase maths attainment. Given the relevance of QRT to the 1stClass@Number 1 intervention and its psychometric properties, we believe this outcome measure is suitable to address the research questions about the impact of 1stClass@Number 1 on maths attainment of pupils.

Secondary outcomes

There are no secondary outcome measures for this trial. The previous trial used Key Stage 1 attainment as a secondary outcome. However, this is not possible in this current trial. In summer 2024 this test will not be mandatory in schools. In addition, the trial schools may have delivered 1stClass@Number 1 to control pupils following the trial period, but before any longer-term attainment outcome measurement would be possible.

Sample size

Table 2: Sample size calculations

		Overall	FSM
Minimum Detectable Effect Size (MDES)		0.14	0.2
Pre-test/ post-test correlations⁸	level 1 (pupil)	0.3	0.3
	level 2 (class)	-	-
	level 3 (school)	-	-
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Total schools		211	211
Number of pupils (per school)	Intervention	4	2
	Control	4	2
	Total	8	4 ⁹
Number of pupils (all schools)	Intervention	844	422
	Control	844	422
	Total	1,688	844

⁸ As this is a pupil-randomised design only the level 1 pre-test/post-test correlation is used in the sample size calculation.

⁹ Assuming 50% of pupils randomised across all schools are eligible for FSM the average number per school will be 4. However, the sample size calculation does not require exactly 4 FSM-eligible pupils at each school: only the overall proportion matters.

While the primary outcome for this evaluation is QRT score amongst all pupils, the sample size was specified so that the evaluation would be powered for a minimum detectable effect size (MDES) of 0.2 in the FSM-eligible pupil subgroup. The MDES of 0.14 amongst all pupils followed from this sample size specification, assuming 50% FSM-eligible pupils. The calculations were performed using the 'PowerUp' software (Dong and Maynard, 2013).

The school and pupil numbers given in Table 2 are those anticipated at the point of randomisation. The recruitment target for this study is 235 schools. Based on previous NFER evaluations we predict approximately 10% of recruited schools will not reach the analysis stage. Since we have taken steps to mitigate attrition post-randomisation, most of these schools are expected to drop out before pupil randomisation¹⁰, meaning that 211 schools will reach randomisation. As 8 pupils will be randomised per school this is 1,688 pupils randomised in total, split evenly between the intervention and control. Assuming 50% of pupils in total across all schools are eligible for FSM (though not necessarily at each individual school), this will be 844 such pupils randomised. Previous NFER evaluations suggest approximately 15% of randomised pupils will not be included in the primary analysis (due to withdrawals, absences on the day of the endpoint test, etc.) This means 1,432 pupils (716 eligible for FSM) will be analysed. Assuming a pre-test/post-test correlation of 0.3, this sample size is sufficient to detect an effect size of 0.14 amongst all pupils (or 0.2 amongst FSM eligible pupils), with a power of 0.8. Some of the assumptions made in this calculation are discussed below.

The pre-test/post-test correlation of 0.3 predicted for this study is noticeably lower than for many EEF trials and is based on the previous evaluation of 1stClass@Number 1. The authors of that evaluation observed an unexpectedly low pre-test/post-test correlation of 0.29 and suggested this was due to the restricted range of maths abilities amongst pupils selected for 1stClass@Number 1, compared to pupils in general. This may have been exacerbated by using the QRT for both pupil screening¹¹ and as a baseline measure. For the current evaluation, we therefore will take the additional step of screening pupils using a different instrument (the SENT-R) to that used for the baseline measurement (the QRT). Whilst we expect this change to improve the pre-test/post-test correlation of the QRT, in the absence of another reliable source a conservative assumption must be made (0.3, similar to the previous evaluation).

This trial is based on an agreed assumption between the evaluator and EHU that half of the pupils randomised will be eligible for FSM. This is higher than in the previous evaluation (where about 33% were eligible for FSM). This decision was taken based on the particular focus for this trial, balanced with the pragmatics of the overall sample size and in particular the number of schools that EHU could work with in a given time period. To that end, additional measures have been taken since the previous evaluation to increase this proportion, as described in the 'participant selection' section above. If the proportion of FSM-eligible pupils

¹⁰ For simplicity's sake all 10% of the schools are treated as dropping out before randomisation in these figures. For the sample size calculation all that matters is that 10% of schools are lost between recruitment and analysis. We will conduct MDES calculations based on actual figures at randomisation and analysis stage, for the report.

¹¹ This was informal screening: QRT score was not an eligibility criterion but was used to inform the teacher's selection of pupils.

transpired to be lower, for example, 40%, this would raise the MDES to 0.22 amongst these pupils.

Randomisation

The eight pupils selected within each school will be randomised in a 1:1 ratio, so that four are assigned to the intervention and four to the control. The pupil selection process is designed so that eight pupils are randomised at each school (see 'Pupil eligibility' above). This is necessary to ensure both a 1:1 randomisation ratio and that the intervention is delivered as intended (in tutoring groups of four – this is a requirement of the intervention implementation). The randomisation will also be stratified by FSM eligibility using the FSM data collected from schools (we will not have access to 'EVERFSM_6_P' from the NPD at this point). Since the impact of 1stClass@Number 1 on FSM pupils is of interest in this trial, stratification will allow a balance of FSM pupils in the intervention and control arms thus maximising statistical power for this group, as a 1:1 randomisation is superior to any other allocation ratio in terms of power. Randomisation will be carried out by NFER statisticians using R Code, which will be stored for reproducibility and transparency. The statistician will not be blinded to group allocation.

Statistical analysis

All statistical analyses will be performed after impact data has been collected for both cohorts and a single dataset has been created including cohort 1 and cohort 2 data. This dataset will include a variable that indicates the cohort to which each pupil belongs to facilitate any additional analyses on this basis.

Primary analysis

The primary outcome is total QRT score measured via tests administered at endpoint by NFER test administrators in schools. This outcome will be the dependent variable in a linear mixed effects regression model, with an independent indicator variable representing the effect of the intervention. Baseline QRT score will be included as a covariate, as will the FSM stratification indicator. Schools will be included as a random effect. For this and all other analyses described below the analyst will not be blinded to treatment allocation.

Secondary analysis

No secondary analysis is planned for this evaluation.

Sub-group analyses

The effect of the intervention for FSM-eligible pupils is of particular interest due to the potentially lower benefit seen for this group in the previous evaluation. Although analysis on FSM-eligible pupils only is most accurately described as a subgroup analysis, this analysis is of central importance to the trial. Sample size calculations were therefore driven by adequately powering the trial for FSM-eligible pupils and results for this subgroup will be emphasised in reporting. Additionally, some analysis that would normally be restricted to the primary analysis sample will also be performed for FSM-eligible pupils (see sections below).

Subgroup analysis will be performed using the 'EVERFSM_6_P' variable from the NPD to define FSM status. The effect of the intervention amongst FSM-eligible pupils will be determined by repeating the primary analysis model (minus the FSM indicator) for this subgroup. Additionally, the differential effect of the intervention for FSM-eligible pupils relative

to non-FSM pupils will be investigated by repeating the primary analysis model with an interaction term between the FSM indicator and treatment indicator added.

Further subgroup analyses will be performed to determine the effect of the intervention for pupils with SEN, defined by converting the 'SENprovisionMajor' variable gathered from the NPD into a binary indicator (SEN versus no SEN). Another subgroup analysis will be performed for pupils with lower SENT-R scores, using the score threshold of 40 and below that was applied as part of the pupil selection process. Each of these subgroup analyses will proceed by first assessing the effect of the intervention in that subgroup and then investigating a differential effect of the intervention using an interaction term, as described above for the FSM subgroup.

If the implementation and process evaluation reveals any notable differences in the implementation of the intervention by cohort, we will perform an additional subgroup analysis to determine the effect of the intervention for pupils in each cohort.

Analysis in the presence of non-compliance

The main compliance measures will be defined at the pupil level and will be based on attendance of tutoring sessions. Specifically, pupils will be considered to have complied with the intervention if:

- (1) they attended at least 30 sessions in total
- (2) they attended at least 5/6 sessions in each of the five topic areas (implying at least 25 sessions attended in total).

A third compliance measure will add the requirement that the pupil's TA attended all 5 of the initial training sessions¹² in some form (catch-up sessions, online or otherwise, are included¹³). This compliance measure therefore uses both the pupil- and TA-level information: both the 5/6 tutoring sessions in each topic by the pupil and 5/5 initial training sessions from the TA are required for the delivery to that pupil to be considered compliant. All three

compliance measures are summarised below:

Table 3: Summary of the six compliance analyses to be performed

Minimum compliance requirements	
Pupil tutoring attendance	TA training attendance
All 30 sessions	No requirement
5/6 sessions in each topic	No requirement
5/6 sessions in each topic	All 5 sessions

For each binary compliance measure a complier average causal effect (CACE) estimate will be obtained using instrumental variable modelling (details will be provided in the statistical analysis plan). The three compliance definitions described above will be modelled for all pupils and given the focus on FSM pupils in this trial, for the FSM-eligible subgroup also, resulting in a total of six models.

Additional analysis

For the primary analysis the choice was made to analyse this multisite trial using a mixed effects model. While this model allows inferences to be made for schools outside the analysed sample, doing so assumes that the effect of the intervention is similar within each school, which may not be the case (Feaster, Mikulich-Gilbertson and Brincks, 2011). To assess the sensitivity of the primary analysis to this assumption, a further modelling approach will be explored. The primary analysis model will be adapted to include an interaction between the school-level random effect and the treatment indicator. This model will be calculated both for

¹² There are six training sessions, but the sixth is not relevant to the TA's ability to teach the five topic areas.

¹³ Note however that the IPE will capture if sessions were catch-up and in what format, in order to describe training attendance.

all pupils and for the FSM subgroup. These additional analyses should be interpreted cautiously, given that the trial is not powered with school by treatment interactions in mind.

Estimation of effect sizes

Effect sizes will be calculated in accordance with the EEF Statistical Analysis Guidance document¹⁴. For the primary analysis the numerator of the effect size will be the difference in means between the intervention and control groups, adjusted for baseline QRT score and stratifiers. The effect size is Hedges' *g* for the primary analysis, so the denominator will be calculated by pooling the standard deviation in the control and intervention groups. A confidence interval for the effect size will be calculated by first constructing a confidence interval for the adjusted mean difference, then dividing the endpoints of the interval by the denominator of Hedges' *g*.

Other analyses will use the same effect size formula as for the primary analysis described above, except for the additional analysis allowing for a school-treatment interaction. This will instead use the effect size and confidence interval formulae provided by (Singh *et al.*, 2022). Details will be provided in the statistical analysis plan.

Implementation and process evaluation (IPE) design

The implementation and process evaluation (IPE) will complement the impact evaluation by exploring how any differences in the implementation of 1stClass@Number 1 may affect the potential impact of the intervention. The IPE design is based on EEF's IPE introductory handbook (and updated in EEF guidance, August 2022¹⁵). It aims to build on the previous evaluations of 1stClass@Number 1 by exploring elements of the logic model that have not been fully explored in previous evaluations, in particular perceived outcomes for pupils and TAs outlined in the logic model, and key moderators and mechanisms relating to pupil selection, pupil motivation and parental engagement. In addition to the compulsory dimensions of fidelity and usual practice, we will also explore any adaptations (e.g., to the training model) that have been made to deliver at scale. Furthermore, because the trial is using a pupil randomised design, we have included a research question to explore how schools felt about this and how well they were able to implement it.

Research questions

IPE RQ1: Who was selected to participate in the intervention and how were they chosen?

Reach:

1.1 Which TAs were selected to deliver the intervention and how were they chosen? What was their previous experience of providing small group support and/or maths support to pupils?

¹⁴<https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1681842352>

¹⁵

https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_guidance.pdf

1.2 How were pupils selected to participate in the intervention? To what extent were the selection criteria and guidance followed (including the focus on selecting FSM pupils) and how did schools feel about this?

Understanding who was selected for the intervention will offer important insights into the extent to which the selection guidance for TAs and pupils was followed by schools and is important context in which to consider the results of the impact evaluation. As part of this research question, we will explore which TAs took part and in particular their perceptions of confidence in and knowledge of maths. This is because choosing someone with successful experience supporting pupils' mathematics is suggested in the TA selection guidance in the 1stClass@Number 1 handbook for schools and because TA mathematical knowledge is a moderator in the intervention logic model. We plan to collect this data using an online survey of TAs conducted at the end of the intervention (see Research Methods section for more details about the survey content).

IPE RQ2: To what extent was fidelity to a) the intended training design, and b) the intended intervention delivery achieved?

Dosage:

2.1 Did TAs and LTs attend all the relevant training sessions? If not, why not?

2.2 Did all pupils receive the intended intervention dosage (i.e., the full 30 sessions in order) (Compliance measure 1)? If not, why not? Were there any differences between FSM and non-FSM pupils?

2.3 To what extent did pupils receive at least 5 out of 6 sessions of each topic, and at least 25 out of 30 sessions in total and did this differ for FSM pupils? (Compliance measure 2)

Fidelity, Adaptation, Quality & Responsiveness:

2.4 Were the training and intervention sessions delivered in accordance with the intervention manual(s)? How variable was this between schools/trainers?

2.5 How did TAs and LTs respond to and engage with the training, and did it provide them with the knowledge, skills, and confidence to deliver the intervention?

2.6 What were LT perceptions of TA delivery quality?

This research question will provide key information about intervention compliance for the impact evaluation and offer insights into the extent to which the training and intervention were implemented with fidelity (i.e., as intended). We are particularly interested in the delivery of the intervention activities which have been changed, adapted and/or re-ordered since the original trial. We will also explore variation in delivery across schools and trainers as we are interested in real-world delivery so would expect some variation in implementation and delivery.

IPE RQ3: How effectively was the intervention implemented in schools, and what supported or hindered successful implementation?

Barriers & Facilitators/Context:

3.1 What were the barriers and facilitators to intervention implementation?

3.2 What contextual factors (e.g., small group) supported or hindered successful intervention delivery (e.g., differences in training/support, TA/LT experience, pupil level factors (including any differences for FSM pupils), school level factors, learning environment, alignment with the curriculum, perceptions of parental engagement)?

3.3 To what extent did the support provided by EHU and LTs facilitate successful intervention delivery?

This research question addresses one of the key aims of this IPE which is to explore some of the moderators outlined in the logic model. From a practical perspective this research question will also provide insights into what is required for successful implementation in schools, which could be used to help shape future implementation guidance.

IPE RQ4: What were the perceived benefits of the intervention for TAs and pupils?

Perceived outcomes (based on the logic model):

4.1 Did the training result in TAs having improved maths knowledge and/or confidence teaching maths and/or making any changes to their practice?

4.2 How did pupils feel they benefitted from the intervention (based on the logic model e.g., maths knowledge, maths confidence, maths self-efficacy, motivation towards maths and mathematical talk)?

4.3 How did TAs feel pupils benefitted from the intervention and did they feel there were any differences between FSM and non-FSM pupils?

4.4 Were there any unexpected/unintended outcomes?

Moderators/Responsiveness:

4.5 Did TAs/LTs feel that any pupils benefitted from the intervention more than others based on pupil characteristics including socio-economic disadvantage? Why/why not?

4.6 How did pupils respond to and engage with the intervention? Why? What facilitated or hindered their engagement?

This research question is the key focus of the IPE for this trial and will investigate the short-term and intermediate outcomes described in the intervention logic model from the perspective of pupils, TAs and LTs, and how these outcomes may vary according to pupil characteristics including FSM eligibility. It will also look at intervention responsiveness for pupils as pupil engagement and motivation are key moderators outlined in the logic model. We also intend to explore how TAs' confidence, knowledge, and ability to tailor in supporting teaching maths may relate to pupil outcomes (see RQ1), by including them as covariates in the impact evaluation regression models.

IPE RQ5: What was business as usual (BAU) and did this change during the intervention?

BAU:

5.1 What was usual practice in intervention schools?

5.2 Did schools make any changes to their planned support during the trial and/or after learning the pupil allocation to intervention/control groups?

Monitoring of control group:

5.3 Did schools provide any additional maths support (apart from 1stClass@Number 1) to either intervention or control pupils during the trial? If so, what maths interventions did intervention/control pupils receive and what was the dosage of these interventions?

5.4 Did TAs feel that they made any changes to their practice as a result of 1st Class@Number 1 which changed the way they worked in the classroom and/or with control pupils?

Understanding BAU is important for determining how the intervention differs from usual practice. In addition, if the intervention or control group receives any other maths interventions during the trial then it could make it challenging to attribute any potential improvements in outcomes. This research question also explores unintended consequences or changes to usual practice that may arise from using a pupil-randomised design for this study, if for example, schools changed the planned support for control pupils after learning their trial allocation.

IPE RQ6: What are the facilitators, barriers, and features of delivering at scale and is the intervention sustainable?

Scale-up:

6.1 What were the facilitators, barriers, and features of training and intervention delivery at scale?

6.2 What are EHU's plans for future training and delivery at scale?

Costs & Sustainability:

6.3 To what extent are schools able to continue resourcing and staffing the intervention going forward? Do they plan to continue using the intervention?

6.4 Would schools be willing and able to pay for the intervention without the subsidy?

6.5 To what extent is the intervention perceived to be cost-effective and offer good value for money?

This is an effectiveness trial, so it is important to understand how well the intervention can be delivered at scale and whether it is sustainable in terms of cost and resource. The previous trial of 1stClass@Number 1 included a full cost evaluation for the intervention and there have been no changes (apart from inflation) to the intervention costs or time required to deliver it since this previous evaluation. To ensure we build on the existing evidence-base, for this evaluation we will be focusing on costs from a sustainability perspective. However, we will update the costs calculated in the previous trial in line with inflation and conduct an uncertainty analysis for staff training costs (see Cost Evaluation Section) to consider different scenarios for how schools may cover staff absence to attend training (e.g., by covering classes internally or getting supply cover).

IPE RQ7: How manageable was it for schools to facilitate the pupil-randomised design?

Context:

7.1 How did schools feel about the randomisation of pupils to groups?

7.2 How well did the randomisation of pupils to groups work in practice? Did it create any challenges (e.g., around managing the group)?

7.3 To what extent did schools adhere to the random allocation of pupils (i.e., not switch children between groups)?

7.4 How well did schools manage the trial and minimise contamination between intervention and control pupils?

1stClass@Number 1 is a pupil-level intervention, making a pupil-randomised design appropriate for this trial. However, a pupil-randomised design creates different considerations to a school-randomised design. Therefore, this final research question is designed to explore schools' views about this type of trial design and the extent to which they were able to successfully implement it to minimise contamination between intervention and comparison pupils. This information will provide important context for the impact evaluation and will also offer insights about schools' views on this type of design which will be relevant for future trials of pupil-level interventions.

Research methods

We are planning to use a variety of different research methods to capture the IPE data which are described below (please also see **Table 3** for a summary of the IPE methods). These methods were confirmed following the project IDEA workshop and set-up meetings which offered the evaluation team the opportunity to speak with EHU in depth about the intervention and the key areas of interest for the IPE.

Based on these discussions, a primary aim of the IPE is to gather evidence about the short-term and intermediate outcomes described in the intervention logic model for both TAs and pupils. Previous evaluations of 1stClass@Number 1 have not explored these outcomes in detail meaning there is limited prior evidence about their relative importance. We therefore decided to take an exploratory approach to understanding these potential outcomes, by looking at them from the perspective of participants. We are particularly interested in the extent to which TAs confidence, knowledge, and ability to tailor the maths sessions are related to pupil outcomes (see RQ1), and the extent to which the training and intervention are perceived to affect such outcomes for TAs (see RQ4).

We are also interested to understand pupils' own views about what they have learned or improved at from taking part in the intervention, and to triangulate these with staff perceptions of pupil outcomes. We felt that qualitative research methods offered us the most appropriate opportunity to explore these potential outcomes in an open way but were also mindful of the age of the children when considering the best ways to capture this information. We felt that as the children would only be 6 or 7 years-old at the end of the intervention and would already be used to taking part in group activities together from the intervention, that they would be most comfortable sharing their views within their existing intervention group rather than one-to-one. We have therefore included pupil focus groups as a key part of the IPE design.

1. TA/LT attendance registers and pupil attendance registers and records

TA and LT training attendance registers will be completed for all participants and be used to monitor training dosage. EHU routinely collect training attendance data, so they will continue to collect this data during the trial and share it with NFER.

Pupil registers and records will be used to monitor attendance at the intervention sessions and will also capture information about any other maths interventions that either intervention or control pupils have taken part in during the trial. EHU typically provide TAs with a paper register they can use to record attendance. NFER will adapt this register to also capture information about any other interventions pupils may have received and create an online system to allow TAs to upload this data and share it with NFER.

We will use the training attendance register data to calculate the average number of training sessions TAs and LTs attend and the proportion of TAs and LTs who attend all their respective training sessions. Pupil attendance records will be used to calculate the average intervention dosage (number of sessions) pupils receive and the percentage of pupils who receive at least 25/30 sessions from 5/6 topics. We will also explore whether there are any differences in intervention dosage for FSM pupils or between the two cohorts. The dosage data will also be used for the trial compliance measures, as described above in the statistical analysis section of the impact evaluation.

2. Trainer observations and reflective interviews

NFER researchers will undertake structured observations and post-session semi-structured reflective interviews (one per trainer) of TA/LT training sessions to monitor the consistency and quality of training delivery across different trainers, as well as to explore perceptions of responsiveness of the TAs/LTs to the training. We note that in some training regions the same trainer may run more than one training group and these training groups may be in the same cohort or in both cohorts. The observations will be spaced throughout the training period and across the two cohorts to allow us to observe each of the 5 training sessions more than once to check consistency between different trainers and between the cohorts.

3. TA observations and reflective interviews

NFER researchers will undertake 10 structured observations of intervention sessions delivered by TAs to check the fidelity and quality of delivery and observe the responsiveness of pupils to the intervention. The observations will be recorded using structured observation proformas and spaced throughout the intervention delivery period to enable us to observe the delivery of each of the 5 intervention topics twice and across the two cohorts to allow us to check for any differences between the cohorts. Content of the observations will include details about the setting in which the intervention is delivered, resources used, the extent to which the full session content was covered (based on the information in the relevant Topic Book), correct use of mathematical language, appropriate adaption to pupils' needs, pupil participation in the session (as an indicator of responsiveness) and perceived outcomes.

Maximum variation sampling will be used to select schools to participate in these observations and reflective interviews in each cohort. This means that sampling will be based on maximum differences in key school characteristics (e.g., geographic region, percentage of FSM pupils, Ofsted rating) to give us the best chance of capturing the widest possible range of experiences. Once we have a list of potential schools based on these characteristics, we will then consider the number of FSM pupils within the intervention groups and where possible prioritise schools with at least two FSM pupils within the intervention groups for the TA observations and reflective interviews as well as the TA/LT interviews and pupil focus groups (see methods 5 and 6). If there are any schools where no FSM pupils have been randomised to the intervention group, these schools will not be selected for the observations or the TA/LT interviews and pupil focus groups as TAs/LTs in these schools will be unable to comment on any potential differences for FSM pupils.

Schools that are selected to participate in the TA observations and reflective interviews will not also be selected to participate in the TA and LT interviews and pupil focus groups, to minimise research burden in schools.

4. TA and LT Surveys

All TAs and LTs who participate in the intervention will be invited to participate in an online survey at the end of the intervention. The surveys will explore specific aspects of a range of IPE dimensions relevant to TAs or LTs (please see Table 4 for an overview of the dimensions covered) and will allow us to capture data from a larger sample of participants than the qualitative data collection methods. The survey questions will primarily be closed questions requiring a single and/or multiple response(s) and will be designed to complement the more detailed information we will obtain from the TA and LT interviews (described below). We envisage that the survey will include questions about TAs' maths qualifications, experience

teaching maths, and experience teaching Year 2 pupils which we plan to score on Likert scales and combine to create a composite measure we can use as a covariate in the regression models described for the impact evaluation. This will allow us to explore any relation between TA maths qualifications and experience and pupil outcomes. Similarly, to explore any relation between TAs' maths confidence and pupil outcomes, we anticipate including adapted versions of some of the subscales from the standardised measure of maths beliefs and confidence by Chen *et al.*, (2014) in the survey. We will then use the total scores from these subscales as covariates in the impact evaluation regression models. The surveys will also contain a marker to indicate which cohort the participant was part of to enable us to take account of any differences by cohort.

Ideally as many TAs and LTs as possible will complete the survey to ensure the sample is comprehensive. To help ensure a high completion rate, the survey will be administered online and the link to the survey (and any subsequent reminders) will be sent directly to the relevant TA or LT as applicable using the contact details obtained at the beginning of the evaluation.

5. TA and LT Interviews

NFER researchers will undertake semi-structured interviews with TAs and LTs which will have a particular focus on perceived outcomes, barriers, and facilitators, scale-up, BAU, and how they felt about taking part in the evaluation. We will also ask about perceptions of parents'/carer's involvement and/or engagement with the 'Special Delivery' activities to explore the relevance of parent/carers involvement to the logic model. Questions will be designed in tandem with those for the surveys to ensure they complement one another, but also provide greater depth on any issues relating to implementation in their particular context.

Interviews will take place at the end of the intervention to allow us to capture TAs' and LTs' experiences from the entire intervention period and allow them to reflect upon any learning or changes for them and/or the pupils they have worked with. We plan to conduct 12 interviews with TAs and 12 interviews with LTs from the same schools. This is to balance data collection efficiency with the probability of achieving sample saturation for qualitative interview data, which typically requires a sample of 9-17 participants (Hennink and Kaiser, 2022). As described in point 3 above, we plan to use maximum variation sampling to select schools to participate in these interviews.

The interviews will take place in the same schools as the pupil focus groups (see below) and will be split across cohorts as this will allow us to triangulate the data from these different perspectives around perceived outcomes and check for any differences between cohorts.

6. Pupil Focus Groups

The focus groups will explore pupils' views about the intervention, and what they think they have learned or got better at from taking part in the intervention, to allow us to explore responsiveness and perceived outcomes from the perspective of pupils. We also plan to ask the children about if and how they have used the 'Special Delivery' activities included as part of the intervention which they can take home and complete with their parent/carers or another family member to explore the relevance of these activities to the intervention logic model.

To ensure that the focus groups are engaging for young children to participate in and encourage them to share their views, we plan to include interactive activities as part of the focus group to help facilitate discussion. We will audio record the focus groups to allow the

researchers to fully engage with the pupils during the session and will also take photographs of the outputs from the interactive activities (with any potentially identifying information removed). We plan to pilot these activities with pupils prior to undertaking the focus groups to ensure they are age appropriate and engaging.

The focus groups will be conducted in the same 12 schools as the TA and LT interviews to maximise data collection efficiency; therefore, the sampling approach will be the same and the focus groups will also be split between cohorts. We are aware that fewer than 12 focus groups is often sufficient for sample saturation, but we are mindful that we are using a heterogeneous sampling approach and may also experience significant attrition if pupils do not receive parent/carer consent to participate and/or are absent on the day of the focus group, and we will not have the opportunity to re-arrange the group. This means that some of the proposed groups may not be able to go ahead. The sample size of 12 therefore takes account of both the sampling approach and the potential for reduced numbers.

Analysis

Qualitative data - Interviews and Focus Groups

Interview and focus group data will be fully transcribed then coded using MAXQDA. Creative outputs from the focus groups will also be coded using MAXQDA. Data from both cohorts relating to the IPE dimensions will primarily be coded deductively based on the IPE dimensions and logic model. This will allow us to collate the evidence for each dimension. We will then undertake a further round of inductive thematic coding within each dimension to identify the key themes which will be described in the final report.

Data relating to perceived outcomes will primarily be coded deductively based on the short-term and intermediate outcomes in the logic model as described above. However, to ensure we capture unexpected and/or unintended outcomes and avoid leading participants' responses we plan to ask about perceived outcomes in an open way in both the focus group and interview topic guides. We therefore also plan to do some inductive coding of data pertaining to perceived outcomes to make sure we capture these unexpected and/or unintended outcomes, which will then be themed as described above.

We plan to use the same coding framework for the observation, interview, and focus group data to allow us to consolidate all the data from relating to each IPE dimension. We are particularly interested to compare the perspectives of different participants from the same schools (i.e., the focus group and interview data). We will therefore label the observations/interviews/focus groups according to the participant group (e.g., LT/TA/pupils) when we import the data into MAXQDA. This will allow us to triangulate perspectives from different groups of participants from the same schools and allow us to use the 'compare cases and groups' function in MAXQDA to look at similarities and differences in responses between different population groups. This approach will give us more detailed insights into the evidence for each element of the logic model. We anticipate that this will be especially important for the data relating to perceived outcomes where this approach will allow us to explore whether TAs' and pupils' own views of what has changed for them match the perceptions of others (i.e., of LTs and/or TAs depending on the participant group).

Quantitative data – surveys, attendance registers, and observations

Quantitative IPE data from the surveys, attendance registers and observations from both cohorts will primarily be analysed together using descriptive statistics. We are yet to design the surveys but expect that most questions will be closed questions requiring a single or multiple responses. Therefore, the descriptive statistics will include percentage response rates to different response options, means and standard deviations depending on the question type.

In addition, to descriptive statistics of survey responses, we plan to use the composite scores of TAs' maths knowledge and confidence (see Research Methods for TA/LT surveys) to explore the potential of TAs' maths knowledge and confidence to moderate intervention outcomes. To do this we will add the relevant composite measure to the regression models used for the impact evaluation to undertake an exploratory analysis of these potential intervention moderators.

Triangulation of qualitative and quantitative data

We plan to design the surveys and interview/focus group topic guides simultaneously. This is so we can ensure that the data we collect for each IPE dimension is captured using the method best suited to the type of data we wish to collect. Therefore, data pertaining to the same IPE dimension may be gathered using more than one method (see Table 4). To integrate these data, we will create an analysis framework outlining how the themes generated from qualitative data (following the deductive coding and additional inductive theming) and each question from the quantitative surveys link to each other and to the IPE dimensions. We will then compare the findings between the different approaches and look to use the qualitative results to help interpret the quantitative results.

Table 4: IPE methods overview

Research methods	Data collection methods	IPE dimension(s)	RQ addressed	Sample size and sampling criteria	Data analysis methods
Surveys	Online surveys for TAs and LTs	Reach, Fidelity, Quality, Adaptation, Perceived Outcomes, Responsiveness	RQ1, RQ2, RQ4	All TAs and LTs will be invited to participate	Descriptive statistics
Interviews	Semi-structured interviews with TAs and LTs	Reach, Fidelity, Quality, Adaptation, Perceived Outcomes, Responsiveness, Barriers & Facilitators, BAU, Scale-up, Context	RQ1, RQ2, RQ3, RQ4, RQ5, RQ6, RQ7	12 LTs and 12 TAs from the same schools ¹⁶ . Maximum variation sampling based on key school characteristics	Deductive coding and inductive coding; thematic analysis
Focus groups	Focus groups with pupils	Perceived outcomes, Responsiveness	RQ4	12 focus groups (in the same 12 schools as the LT and TA interviews)	Deductive and inductive coding; thematic analysis
Registers	Training attendance registers (TAs and LTs) Intervention attendance registers & pupil record (pupils)	Dosage / compliance, Monitoring of control group	RQ2, RQ5	All TAs and LTs All pupils	Frequency counts; descriptive statistics
Observations and reflective interviews	Structured observations and reflective interviews	Fidelity, Quality, Adaptation, Responsiveness, Perceived Outcomes	RQ2	16 x Trainer observations (one per trainer) 10 x TA observations (2 per topic) selected using maximum variation sampling based on key school characteristics.	Frequency counts; deductive coding; thematic analysis

¹⁶Note that the 10 TA observations and reflective interviews will not be conducted in the same schools as the 12 TA and LT interviews and pupil focus groups to allow us to capture perspectives from different schools.

Cost evaluation design

A full cost evaluation was conducted as part of the previous trial of 1stClass@Number 1 (Nunes *et al.*, 2018). There have been no changes to the costs of the programme in terms of the programme model, resources or teacher time required since this evaluation other than adjustments for inflation. It was therefore agreed that to add most value to the existing evidence base, we would focus on costs in this trial from the perspective of sustainability and perceived value for money as part of the IPE (see Section on IPE design), as opposed to repeating the full cost evaluation.

However, we recognise that costs have increased in relation to inflation and that some of the EEF guidance around cost evaluations¹⁷ has been updated since the previous trial. For these reasons, we have agreed that for this trial we will do the following:

- **Adjust the costs calculated for the previous trial to take account of inflation.**

To do this we will primarily use the costs calculated by the previous trial and adjust them for inflation (as described in EEF Cost Evaluation Guidance, 2023). Where the new cost is known (e.g., the cost of the training package to schools), we will use the known cost in the revised calculation.

- **Conduct a sensitivity analysis for personnel training costs.**

The original trial did not cost for supply cover for the TA or LT to attend training. We therefore propose to conduct a sensitivity analysis as per Principle 6 of the EEF Cost Evaluation Guidance (EEF, 2023) to take account of whether schools chose to cover staff absence for training internally or bring in supply staff.

We believe it is very unlikely that schools would bring in supply cover for the TA, so for the purposes of the cost evaluation we will assume that TA absence was covered internally. (We will, however, check this assumption with schools, and adjust our approach to updating the cost calculations if required.) We will use the LT survey to understand if supply cover was used to cover the LT's training time, and if so, to estimate the cost of this cover. Based on this survey data, we will use the prevalence of internal vs external cover to determine which option we present as the typical cost of the programme for schools (i.e., if most schools covered the LT internally, we would use this cost in the primary cost analysis), with the alternative option presented in the sensitivity analysis to show the potential range of costs to schools depending on how they choose to cover the LT training time.

¹⁷https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1684358910

Ethics and registration

This evaluation will be conducted in accordance with NFER's Code of Practice, available at [NFER Code of Practice](#). All of NFER's projects abide by its Code of Practice, which is in line with the Codes of Practice from BERA (the British Educational Research Association), MRA (the Market Research Association) and SRA (the Social Research Association), among others. NFER is committed to the highest ethical standards in all of its activities and ethical considerations are embedded in its detailed quality assurance processes.

Each participating school's headteacher will provide their agreement to participate in the trial by signing the Memorandum of Understanding (MoU) that outlines the responsibilities of all parties involved in the trial. NFER will share a parent letter and withdrawal form with schools to be sent to parents/carers of all pupils that schools intend to nominate for screening. Through the withdrawal form, parents/carers will have the opportunity to withdraw their child from the evaluation and associated data processing at any stage of the trial.

A separate opt-in consent process will be used for the pupil focus groups and will only apply to those selected to participate. Given that pupils participating in this study are only 6 to 7-years-old, we cannot assume that all pupils will have the capacity to provide fully informed consent to participate. In addition, as the focus groups involve audio recordings, it is especially important to ensure that parents/carers have the option to specifically consent to their child participating in this evaluation activity. We will therefore provide parents/carers with a written information sheet about the focus groups which will contain full details about the focus group and what their child will be asked to do. Parents/carer will then be asked to provide written opt-in consent of their willingness for their child to be invited to participate in the focus group, by returning a consent form to the school, who will then pass this information on to the research team.

Pupil participation in the focus groups is voluntary, therefore even if a parent/carer has given consent for their child to participate, their child can still choose not to take part. Age-appropriate information about the focus groups will be provided to pupils at the same time as parents/carers receive information about the focus groups to allow them to discuss participation together. The researchers will also read this information to pupils at the beginning of the focus group to ensure pupils understand it and have the chance to ask any questions. If at this point a pupil decides that they would prefer not to participate, then they will be able to return to their class. Prior to beginning the focus group, the researchers will agree some ground rules for the group with the pupils and have a discussion with them about the types of scenarios in which we would need to break confidentiality, to ensure they fully understand what this means.

Data protection

All data gathered during the trial will be held in accordance with the data protection framework created by the Data Protection Act 2018 and the General Data Protection Regulation 2016/679 and will be treated in the strictest confidence by the NFER, EHU and EEF. No individual or school will be identified in any report.

NFER and EHU are joint data controllers for the duration of this trial.

The legal basis for processing personal data is covered by: GDPR Article 6 (1) (f) which states that 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'. We have carried out a legitimate interest assessment, which demonstrates that the evaluation fulfils one of NFER's core business purposes (undertaking research, evaluation and information activities) and it has broader societal benefits. Therefore, it is in our legitimate interest to process and analyse personal data for the administration of this RCT.

NFER and EHU have signed a Sharing Agreement that will govern the collection and sharing of personal data during this trial. This agreement includes a description of the nature of the data being collected and how it will be shared, stored, protected and reported by each party. In addition, EHU will provide a memorandum of understanding to schools, explaining the nature of the data being requested of schools, teachers and pupils, how it will be collected, and how it will be passed to and shared with NFER. Two separate Privacy Notices are available: one for schools¹⁸ and another one for parents¹⁹.

For the purposes of the trial, EHU will collect names, role and contact details of a key contact person and the person signing the MoU when schools are recruited. They will share these data with NFER, who will then contact the key project contact person at participating schools to initiate data collection for the trial. NFER will ask participating schools to nominate 12 – 16 eligible pupils for screening in addition to one TA and link teacher who will support with intervention delivery. NFER will share the names, contact details and job roles of the TAs and link teachers with EHU to enable them to coordinate the training sessions. EHU will collect attendance data for TAs and link teachers at each training session and will share these data with NFER to create compliance measures that will be linked to pupil-level data. EHU will also share the names and contact details of 1stClass@Number 1 trainers with NFER, so that NFER can contact them to attend and observe a sample of training sessions. All personal data will be shared via secure, password-protected data sharing portals.

NFER will also collect pupil data from schools including names, date of birth, Unique Pupil Number (UPN) and FSM eligibility status for 12 – 16 eligible pupils who will subsequently be screened using the SENT-R. The SENT-R and baseline QRT scores will also be collected for these pupils. In addition to these data, for the 8 pupils finally selected to participate in the trial, NFER will also collect the endpoint QRT scores, details of any other interventions received, class name and attendance at intervention sessions (if allocated to the intervention group).

¹⁸ [eeqn_privacy_notice_for_schools.pdf \(nfer.ac.uk\)](https://www.nfer.ac.uk/privacy-notice-for-schools.pdf)

¹⁹ [eeqn_privacy_notice_for_parents.pdf \(nfer.ac.uk\)](https://www.nfer.ac.uk/privacy-notice-for-parents.pdf)

For these pupils, background data including, gender, FSM eligibility status and SEN status will be collected from the National Pupil Database (NPD). To obtain the information from the NPD, NFER will provide the Data Sharing Team at the DfE with the names of the pupils, their dates of birth and UPNs, allowing a match to NPD.

As part of the implementation and process evaluation (IPE), NFER will conduct online surveys of TAs and link teachers, observations of training sessions, interviews with trainers, observation of lesson delivery by TAs and interviews with TAs and link teachers. NFER may share summaries of pseudonymised data from TAs' or link teachers' interviews and aggregated anonymised data from the TA and link teacher surveys with EHU to support their research. A small number of schools will also be invited to participate in pupil focus groups. All NFER staff visiting schools will have up-to-date DBS checks. All data gathered during interviews will be stored securely. No names of individuals will be used in any report arising from this work.

Within three months of the end of project, NFER will send school and pupil data to EEF's data archive partner. At this point, EEF's data archive partner will keep a copy of the data and EEF will become the Data Controller. NFER will retain personal data for one year after report publication in case there are any queries about the report. One year after the report publication, all personal data will be securely deleted.

Personnel

Name	Organisation	Role and Responsibilities
Pippa Lord	NFER	Project Director – responsible for overall delivery of the trial, IPE Director
Aarti Sahasranaman	NFER	Trial Manager – day-to-day management of the trial and delivery of the trial design
Emma Moore	NFER	IPE Lead – design and delivery of the IPE
Eleanor Bradley	NFER	IPE Researcher – IPE data collection
Chris Morton	NFER	Statistician
Kathryn Hurd	NFER	Research Operations Lead - overall data collection and school communications strategy
Jishi Jose	NFER	Project Manager – day-to-day operations including preparation of recruitment documents, coordinating data collection and point of contact for schools participating in the trial
Lydia Wallis	NFER	Operations Researcher – support Project Manager in delivery of data collection and school communications strategy
Yvonne Panteli	EHU	Project Manager – day-to-day operations including coordination of school recruitment, set up of programme delivery.

Lisa Leach	EHU	Every Child Counts Development Lead – school recruitment and intervention delivery plus school visits.
Linda Slater	EHU	Every Child Counts Development Lead – school recruitment and intervention delivery plus school visits.
Jonathan Glazzard	EHU	Contributed at project set-up, will not be involved with project from July 2023
Jane Moore	EHU	Dean of Faculty of Education – responsible for the overall set up and delivery for the trial
Peter Hick	EHU	Principal Investigator for Edge Hill – responsibility for providing academic oversight on behalf of the University and leading on a parallel project on Teaching Assistant professional development.

Risks

S. No.	Risk	Risk Assessment		Mitigation/Counter Measures/Contingencies
		Likelihood	Impact	
1	Insufficient schools recruited by EH (Target: 235)	May Happen	Significant	Weekly updates from EHU on recruitment; monthly check-in meetings with EH and EEF NFER to support recruitment (with added budget) Extend recruitment window, consider delaying delivery, consider delivery over 2 cohorts. Consider powering the trial for all pupils rather than FSM pupils (145 schools with MDES = 0.18 OR 117 schools with MDES = 0.20); consider the value of the results in the context of a meta-analyses.
2	Schools are unable to nominate at least 12 pupils for screening	May Happen	Moderate	Weekly updates from EHU on recruitment to include data from EoI indicating nomination capacity of schools. Provide nomination guidance early to schools to identify appropriate pupils for screening. Schools allowed to nominate a minimum of 8 pupils (plus an additional 2 pupils to accommodate any change in circumstance of pupils) if 12 is not possible (included in pupil nomination guidance). Schools

				are requested to discuss this with NFER and/or EHU. Schools that cannot nominate a minimum of 8 pupils cannot participate in the trial. Additional schools may need to be recruited to achieve sample size.
3	Fewer than 8 pupils from each school eligible for selection based on Sandwell test scores	May Happen	Significant	Schools asked to nominate 12 - 16 pupils who will then be screened using the Sandwell test to confirm eligibility and select 8 pupils to the trial. Relax the Sandwell cut-off score criterion and allow pupils close to but above the cut-off score to be selected to the trial. For example, if only 4 pupils are below the cut-off score, identify the next 4 pupils who are closest to the cut-off score and select these 8 pupils to the trial. NB: The upper-most raw score for a child to be chosen for 1stClass @Number 1 is 49, which equates to a number age of 6:10.
4	Insufficient FSM pupils take part in the trial / schools do not follow the selection guidance regarding FSM	May Happen	Moderate	All recruitment documents to highlight focus of study on disadvantaged pupils; pupil nomination guidance to highlight that schools should nominate FSM-eligible pupils who meet eligibility criteria; monitor Y1 and school-level FSM % during recruitment, and FSM % among nominated pupils when collecting pupil data. Focus recruitment on schools with higher proportion of FSM pupils, including re-target in latter phase of recruitment; revisit FSM pupil nomination guidance. Consider recruiting more schools (with increased budget) (e.g. if 40% are FSM, pre-post of 0.3, requires 294 schools. If we assume an improved pre-post correlation of 0.4, 40% FSM requires 272 schools) Consider recruiting sufficient schools to be powered for all pupils, and consider FSM

				results in the context of their value in meta-analyses. Interaction term analysis to assess impact of FSM
5	Control pupils receive the intervention	Very unlikely	Moderate	Because this is a pupil-randomised trial, there is a risk of contamination. However, this risk is deemed very low, given that the TA will be delivering the intervention to pupils in a small group outside of the classroom. TAs will have a register for the pupils in the intervention group. Control pupils will remain in their normal classroom setting. Schools will be specifically asked not to switch pupils' allocation. It is therefore unlikely that control group pupils will receive any aspects of the intervention. TAs will be provided with a Dos and Don'ts guidance document to alert them to the possibility of contamination. Schools will also be alerted of the possibility of delivering 1stClass@Number 1 to control pupils after endpoint testing has been completed for intervention pupils.

Timeline

Table 5: Timeline

Dates	Activity (Cohort 1)	Activity (Cohort 2)	NFER Staff responsible/ leading
Dec '22 – Feb '23	Project set-up Complete project set-up, finalise recruitment documents		Pippa Lord, Aarti Sahasranaman
Mar '23 – Sep '23	Recruitment and pupil data collection Recruitment of cohort 1 and cohort 2 schools by EHU – recruitment of cohort 1 schools to be completed by end of summer term, recruitment of cohort 2 schools to continue until end of third week of September. EHU share school data for cohort 1 and 2 schools with NFER, NFER contact schools to collect pupil data for nominated pupils and TA and link teacher data. NFER prioritise collection of pupil data from cohort 1 schools before end of summer term. Collection of pupil data from cohort 2 schools ongoing, to be completed in the second week of October.		Kathryn Hurd, Jishi Jose
	Study Protocol Submit draft protocol, revise based on feedback and publish study protocol		Pippa Lord, Aarti Sahasranaman
Jul '23 – Aug '23	Preparation of assessment materials and administration guidance for cohort 1 schools NFER prepare Sandwell Early Numeracy Test – Revised (SENT-R) and baseline QRT test booklets for nominated pupils in cohort 1, develop guidance for teachers to administer SENT-R and QRT, develop guidance for processing SENT-R scores		Jishi Jose, Aarti Sahasranaman
	Start design of IPE instruments		Emma Moore, Eleanor Bradley
Sep '23	Sandwell screening and baseline QRT and selection of pupils to trial in cohort 1 schools SENT-R and baseline QRT administered in cohort 1 schools,	Complete pupil data collection from cohort 2 schools	Kathryn Hurd, Jishi Jose

	test booklets sent to NFER for marking/processing, NFER send Sandwell score report to schools recommending pupils to be selected to the trial, schools confirm final 8 pupils for trial		
	Complete design of IPE instruments		Emma Moore, Eleanor Bradley
Oct '23	Randomisation of pupils in cohort 1 schools Randomisation of selected pupils in cohort 1 schools	Preparation of assessment materials for cohort 2 schools NFER prepare Sandwell Early Numeracy Test – Revised (SENT-R) and baseline QRT test booklets for nominated pupils in cohort 2	Chris Morton, Aarti Sahasranaman
	Notify cohort 1 schools of study arm to which each pupil is assigned		Kathryn Hurd, Jishi Jose
	Complete marking and QA of baseline QRT data from cohort 1 schools		Kathryn Hurd, Jishi Jose
	Submit draft Statistical Analysis Plan (SAP)		Chris Morton, Aarti Sahasranaman
Nov – Dec '23	See below	Sandwell screening and baseline QRT and selection of pupils to trial in cohort 2 schools SENT-R and baseline QRT administered in cohort 2 schools, test booklets sent to NFER for marking/processing, NFER send Sandwell score report to schools recommending pupils to be selected to the trial, schools confirm final 8 pupils for trial	Kathryn Hurd, Jishi Jose
Nov '23 – Mar '24	Training and intervention delivery and data collection for cohort 1 schools 1 st Class@Number 1 training and intervention delivery for cohort 1 schools	See below	EHU, schools
	IPE data collection in cohort 1 schools– trainer and TA observations and reflective interviews		Emma Moore, Eleanor Bradley

	Submit NPD data application (Jan '24)		Chris Morton, Aarti Sahasranaman
	QRT endpoint testing in cohort 1 schools by NFER test administrators, TA/ LT surveys (Mar '24)	Complete marking and QA of baseline QRT data from cohort 2 schools	Kathryn Hurd, Jishi Jose
	Obtain pupil attendance/other intervention records from cohort 1 schools (Mar '24)		Kathryn Hurd, Jishi Jose
	Case study visits for cohort 1 schools (Mar '24)		Emma Moore, Eleanor Bradley
Jan - Jun '24	See above and below	Training and intervention delivery and data collection for cohort 2 schools 1 st Class@Number 1 training and intervention delivery for cohort 2 schools	EHU, schools
		IPE data collection in cohort 2 schools– trainer and TA observations and reflective interviews	Emma Moore, Eleanor Bradley
		QRT endpoint testing in cohort 2 schools by NFER test administrators, TA/ LT surveys (May/Jun '24)	Kathryn Hurd, Jishi Jose
	Obtain training attendance data for both cohorts from EHU (Jun '24)		Kathryn Hurd, Jishi Jose
		Obtain pupil attendance/other intervention records from cohort 2 schools (Jun '24)	Kathryn Hurd, Jishi Jose
		Case study visits for cohort 2 schools (May/Jun '24)	Emma Moore, Eleanor Bradley
	Publish Statistical Analysis Plan (SAP)		Chris Morton, Aarti Sahasranaman
Apr – May '24	Complete marking and QA of endpoint QRT data from cohort 1 schools		Kathryn Hurd, Daniel Finn
Jun – Jul '24		Complete marking and QA of endpoint QRT data from cohort 2 schools	Kathryn Hurd, Daniel Finn
Jul – Aug '24	QA and analysis of IPE data		Emma Moore, Chris Morton

Sep – Nov '24	Access data on NPD, complete primary and additional analyses	Chris Morton, Aarti Sahasranaman
Jan '25	Submit first draft of report	Pippa Lord, Aarti Sahasranaman
Jun/Jul '25	Final report published	Aarti Sahasranaman
Sep '25	Submit data to EEF archive	Chris Morton, Aarti Sahasranaman

References

- 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), pp. 367–377. Available at: <https://doi.org/10.1007/s10643-013-0619-0>.
- Dong, N. and Maynard, R. (2013) 'PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies', *Journal of Research on Educational Effectiveness*, 6(1), pp. 24–67. Available at: <https://doi.org/10.1080/19345747.2012.673143>.
- Education Endowment Foundation (2023) *Cost evaluation guidance for EEF evaluations*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1684358910 (Accessed: 30 May 2023).
- Feaster, D.J., Mikulich-Gilbertson, S. and Brincks, A.M. (2011) 'Modeling Site Effects in the Design and Analysis of Multi-site Trials', *The American Journal of Drug and Alcohol Abuse*, 37(5), pp. 383–391. Available at: <https://doi.org/10.3109/00952990.2011.600386>.
- Hennink, M. and Kaiser, B.N. (2022) 'Sample sizes for saturation in qualitative research: A systematic review of empirical tests', *Social Science & Medicine* (1982), 292(2022), p. 114523. Available at: <https://doi.org/10.1016/j.socscimed.2021.114523>.
- Nunes, T., Barros, R., Evangelou, M., Strand, S., Mathers, S. and Sanders-Ellis, D. (2018) *1stClass@Number: Evaluation report and executive summary*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/projects/1stClass@Number_evaluation_report.pdf?v=1630925665.
- Nunes, T., Bryant, P., Evans, D. and Barros, R. (2015) 'Assessing quantitative reasoning in young children', *Mathematical Thinking and Learning*, 17(2–3), pp. 178–196. Available at: <https://doi.org/10.1080/10986065.2015.1016815>.
- Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A. and Carraher, J. (2007) 'The contribution of logical reasoning to the learning of mathematics in primary school', *British Journal of Developmental Psychology*, 25(1), pp. 147–166. Available at: <https://doi.org/10.1348/026151006X153127>.
- Singh, A., Uwimpuhwe, G., Li, M., Einbeck, J., Higgins, S. and Kasim, A. (2022) 'Multisite educational trials: estimating the effect size and its confidence intervals', *International Journal of Research & Method in Education*, 45(1), pp. 18–38. Available at: <https://doi.org/10.1080/1743727X.2021.1882416>.

Appendix A: Changes since the previous EEF evaluation

Appendix Table 1: Changes since the previous evaluation

Feature		Previous effectiveness trial	Current effectiveness trial
Intervention	Intervention content	One of the components of the intervention helps schools enhance or develop their approach to parental engagement in their child's mathematical learning. Children were given a mathematics activity called 'Special Delivery' every fortnight to take home and practice what they had learned. Parents were asked to help children do the activity and talk to them about it. A Post Office theme was used throughout the programme.	The formative evaluation by the BIT suggested that the Post Office theme could be less beneficial for disadvantaged pupils and recommended updating it to a new context that does not disadvantage FSM pupils. ECC have now updated the content of the 'Special Delivery' to reflect a Parcel Delivery theme.
	Delivery model	No change	
	Intervention duration	No change	
Evaluation	Eligibility criteria	1stClass@Number 1 is for Year 2 pupils who have fallen behind their peers and need more support at the level of the Year 1 curriculum. They may typically: • be able to count forwards <u>and backwards</u> in ones from 10 • have some knowledge of number facts • be able to perform simple addition and subtraction calculations using counting all and counting out approaches • be able to read and write numbers but lack secure understanding of their magnitude and quantity • struggle with <u>mental calculation strategies and</u> mathematical vocabulary • lack confidence in mathematics and be reluctant to talk about their mathematical learning. Teachers were also advised to nominate for the same group children who can 'gel' and learn together and were not participating in another intervention, in order to avoid overload.	One of the main areas of focus in this trial is the pupil selection mechanism. In order to identify pupils eligible to participate in this trial, we will follow a two-step process. In step 1, schools will be asked to nominate 12 – 16 pupils (if possible) that they feel are eligible to receive 1CN1 (note, 10 is the minimum they can nominate, i.e. 8 plus a further 2 to mitigate against any attrition). In step 2, these nominated pupils will be objectively screened using the Sandwell Early Numeracy Test (Revised) (SENT-R) to identify the final 8 pupils who will be selected to participate in this trial. The eligibility criteria for pupils to be nominated to the trial are similar to the previous trial. The changes are underlined. Suitable pupils may typically: • be able to count forwards in ones to 10 • have some knowledge of number facts <u>and also have some understanding of the composition of number within 10</u> • be able to perform simple addition and subtraction calculations using counting all and counting out approaches

		The intervention team gave guidance to schools on how they might take into account the pupils' scores in the QRT pre-test as part of the information to be considered when they nominated pupils for participation in the project.	• be able to read and write numbers but lack secure understanding of their magnitude and quantity • struggle with mathematical vocabulary • lack confidence in mathematics and be reluctant to talk about their mathematical learning. Additionally, because the trial has a focus on disadvantaged pupils, schools will be encouraged to nominate FSM-eligible pupils who meet the above criteria. To avoid overload, pupils participating in other maths interventions are not eligible for this trial. Because of the pupil-level randomisation, schools were not specifically asked to select pupils who can 'gel' and learn together. In the next step, the SENT-R will be administered 1:1 to all nominated pupils. NFER will share a customised report with each school rank ordering pupils based on their raw SENT-R scores and recommending 8 pupils for selection to the trial. Schools will then be asked to confirm the final 8 participants to be selected to the trial.. In recommending pupils for selection to the trial, we will first prioritise the selection of FSM-eligible pupils with raw SENT-R scores of 49 or below. While pupils scoring 40 or below are considered to be around 12 months behind age-related expectations, in agreement with EHU we are also allowing the selection of pupils with a raw score between 41 – 49, where a raw score of 49 equates to a number age of 6:10 (this is part of usual practice that EHU experience). Once FSM-eligible pupils have been recommended, we will randomly choose non-FSM pupils with raw scores of 40 and below to reach the target of 8 pupils for recommendation to the trial. If this target is not achieved by choosing pupils with scores of 40 or below, we will then randomly choose non-FSM pupils with a raw score of 41 - 49 in order to reach the required number of 8 pupils. In the event that a school does not have 8 pupils who fall below the score of 49 on the SENT-R, we will then randomly choose pupils with scores above 49 to reach the target of 8 pupils (this is preferable to randomising fewer than 8 pupils in a school, further details below). Schools will be asked to briefly confirm the recommended shortlist, and also given the option of replacing recommended non-FSM pupils with other pupils who have not been recommended by us.
--	--	---	---

	Level of randomisation	School-level randomisation	Pupil-level randomisation within schools
	Outcomes and baseline	No change in baseline and primary outcome. The secondary outcome was Key Stage 1 (KS1) maths. KS1 maths results were obtained through the NPD using the variable KS1_MATH_OUTCOME.	We were unable to use KS1 maths as a secondary outcome in this study for two reasons. In summer 2024, this test will not be mandatory in schools. Furthermore, the trial schools may have delivered 1 st Class@Number 1 to control pupils following the trial period, but before any longer-term attainment outcome measurement would be possible.
	Control condition	Control condition remains business-as-usual	Although the control condition remains the same, because this is a pupil-randomised trial, there is a risk of contamination that was not present in the previous trial. However, this risk is deemed low, given that the TA will be delivering the intervention to pupils in a small group outside of the classroom. TAs will have a register for the pupils in the intervention group. Control pupils will remain in their normal classroom setting. Schools will be specifically asked not to switch pupils' allocation. It is therefore unlikely that control group pupils will receive any aspects of the intervention.