Education
Endowment
Foundation

Counting Collections

Evaluation Report

July 2025

Martin Culliney, Karen Daniels, and Joanne Robson

Sheffield Hallam University
Knowledge Applied

Sheffield Institute of Education

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

We do this by:

- **Summarising evidence**—reviewing the best available evidence on teaching and learning and presenting it in an accessible way;
- **Finding new evidence**—funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socioeconomically disadvantaged backgrounds; and
- **Putting evidence to use**—supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

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

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

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

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The evaluation team are based at the Sheffield Institute of Education, Sheffield Hallam University (SHU). As members of the EEF evaluator panel for more than ten years they have completed numerous randomised controlled trials and other studies for the EEF and organisations including government departments, charities, and other policymakers.

The project email address is Counting-Collections@shu.ac.uk

Acknowledgements

As well as the report authors, the project team included Catherine Hathaway, Fiona Storey, and Jacqueline Launders who undertook fieldwork as part of the implementation and process evaluation. Nick Speed, Tom Kidston, Freya Wallace, and Linda Bray all provided valuable administrative support. The assessment administrators, led by Katie Morris and Rachel Brown at Qa Research, collected baseline and outcome data in participating schools and all made a crucial contribution to the study. Finally, the evaluators would like to thank the Counting Collections development team at the University of Nottingham, the teachers and children who took part in the evaluation, and colleagues at the EEF who were involved in the trial.

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

Executive summary

The project

Counting Collections is a hands-on approach to developing early number sense (understanding of number and quantity) in reception aged children (4-5 years old) which supports counting, subitising, comparing numbers, and composition of numbers. It involves pupils using containers of objects (manipulatives) to find out how many are in the collection. A structured routine for counting is taught in four parts: choose, strategise, count, record. This routine represents one of the core components of the programme. The intervention also involves adding a Counting Collections area, known as a Counting Library, to the classroom which pupils can continue to access during continuous provision. One teacher from each school attends Continuing Professional Development (CPD) training provided by the developer. The first training session is delivered in person with the four remaining sessions taking place online. Teachers deliver the 30-minute intervention to the whole class, once a week for 20 weeks across one school year. The Counting Collections programme is developed by a team led by Dr. Catherine Gripton, Associate Professor in the School of Education at the University of Nottingham.

The evaluation was a two-armed randomised controlled efficacy trial, with randomisation taking place at school level. A total of 180 schools were recruited to the study with one reception class participating form each school. The implementation and process evaluation involved interviews with school staff, observations of lessons and teacher training sessions, and a post-intervention teacher survey. The trial took place between February 2023 and July 2024.

The study was funded by the Education Endowment Foundation (EEF) through the Department for Education's Accelerator Fund.

Table 1: Key conclusions

Key conclusions	
1.	Children in Counting Collections schools made the equivalent of one month's progress in number attainment, on average, compared to children in the control schools. This result has a moderate to high security rating.
2.	Among children eligible for free school meals, those in Counting Collections schools made no additional months' progress in numeracy, on average, compared to children eligible for Free School Meals in other schools. These results may have lower security than the overall findings because of the smaller number of pupils.
3.	Teachers who had taken part in Counting Collections were more likely to report greater confidence and 'major improvements' in teaching and supporting pupils with early number interactions than teachers in control schools.
4.	Participating teachers reported that the programme improved pupil collaborative working, recording, mathematical thinking, and mathematical talk. They also reported increased engagement, concentration, motivation, number confidence, and enjoyment of mathematics.
5.	Observational findings indicated that most pupils were engaged in high quality number experiences supported by adult-child interactions, though some teachers found the four-part routine challenging and suggested more guidance for extending higher attainers and supporting lower attainers.

EEF security rating

These findings have a moderate to high security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools. The trial was a well-designed two-armed randomised controlled trial. The trial was well-powered. The pupils in Counting Collection schools were similar to those in the comparison schools in terms of prior attainment. However, the security of the trial was reduced by the following factors: 24% of the pupils who started the trial were not included in the final analysis because their school did not provide pupil lists, the assessment data was lost during the collection process, or pupils were absent on testing days.

Additional findings

Pupils in Counting Collections schools made, on average, one additional month's progress than those in the control group equivalent. This is our best estimate of impact, which has a moderate to high security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes no additional progress and a positive effect of up to two months of additional progress.

Compliance with the intervention was high. This was measured through three indicators: attendance at training sessions, presence of intervention resources in classrooms, and delivery of intervention sessions. Of the 87 intervention schools to complete the trial, 79 were fully compliant. For those that were not, it was because the teacher did not attend all the training sessions. Analysis indicates that different levels of compliance did not impact the trial results.

The trial suffered from attrition with an estimated 24% of pupils randomised not completing outcome assessments. This was based on projected pupil numbers as full pupil lists were submitted post-randomisation. Eight of the 180 schools randomised (two intervention and six control) did not take part in baseline testing. The baseline test papers from two schools were lost in the post, and one school withdrew from the study during delivery. From the 169 schools taking part in outcome testing, all other attrition was at pupil level. Sensitivity analysis shows that the findings appear robust after multiple imputation for missing outcome data.

Planned activities or inputs specified by the logic model were observed during the trial, as evidenced through training observations, the teacher survey, and fieldwork visits. Outputs reported include improved teacher knowledge and confidence in planning, teaching and interacting with pupils, and the use of learning trajectories to support progression.

The whole-class session format integrated successfully into existing classroom practice in maths teaching. Over the last ten years schools have been encouraged to use a mathematics mastery approach (Boylan, 2018) and mathematics schemes in primary school have seen a resurgence (Turvill, 2021). Most intervention schools reported that the whole-class approach involving the four-part routine complemented and strengthened the mathematics provision already used in class.

Most teachers also felt that the programme needed teaching assistant (TA) support. Schools that had a TA often deployed them to work with pupils who were struggling with basic counting. Teachers expressed concern about how to support those pupils through the four parts of the programme routine. Future iterations of the programme may consider further training TAs to support delivery. It would also be worth exploring how the programme can be tailored to support pupils who are less secure in basic counting.

Cost

The average cost of Counting Collections for one school was around £700—just over £11 per pupil per year when averaged over three years for a class of 20 pupils.

Impact

Table 2: Summary of impact on primary outcome(s)

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P Value	EEF cost rating
SENT-R (numeracy) all pupils	0.08 (-0.02; 0.18)	1	🔒🔒🔒🔓🔓	2741	0.097	£££££
SENT-R (numeracy) FSM pupils	0.03 (-0.12; 0.17)	0	N/A	650	0.709	£££££

Introduction

Background

Pupils start school with different experiences of number. This affects how they access opportunities to learn number, through play and through teacher-directed activities, creating a gap in early number attainment. Research suggests that early number skills are highly correlated with mathematics outcomes (Schneider et al., 2017) and a strong predictor of later achievement in mathematics (Göbel et al., 2014; Sasanguie et al., 2012). Number sense is crucial to mathematics attainment and a powerful predictor of later mathematics outcomes (Jordan et al., 2010). There is evidence that children from low-income families typically have lower mathematics attainment than their peers and that this gap is already present when children begin formal schooling (Short and Mclean, 2023). This highlights the importance of building on the mathematical potential of young learners (Bisaillon, 2023). Systematic and regular counting experiences are shown to help in addressing any gaps that may exist from preschool number experiences and instructional approaches that capture and sustain children's interest are seen as being essential in changing the mathematics learning trajectories (Jacobi-Vessels et al., 2016). Effective instructional approaches are of particular importance for children from socioeconomically disadvantaged backgrounds (Denton and West, 2002).

Counting Collections is a hands-on early mathematics approach to number sense (understanding of number and quantity) and involves a teacher-led sequenced structure. It supports counting, subitising, comparing numbers, and composition of numbers. Pupils work in pairs with containers of objects (manipulatives), following a four-part sequence where they choose, strategise, count, and record how many items are in the collection. The intervention includes opportunities for pupils to engage in number interactions with both teachers and peers. Effective learning in the early years comes through quality interactions with adults, including interactions during planned and child-initiated play and activities (Ofsted, 2022). The adult's understanding of intentional mathematical learning through number interactions is therefore pivotal to pupil progress. This requires the practitioner to have confidence and knowledge of mathematics as an interconnected subject to 'ensure every interaction with a child is an opportunity for teaching and learning' (Ofsted, 2021). It is important that timely interactions, modelling, questioning, and reflective commentary are used to direct the young learner, requiring good knowledge of the complexities of number development. Number sense encompasses several skills including verbal counting, knowing the number symbols, recognising quantities, discerning number patterns, comparing numerical magnitudes, and manipulating quantities (Raghuba and Barnes, 2017). It involves an understanding of the value and relationships between numbers and is a foundation for all higher-level mathematics (Feikes and Schwingendorf, 2008). The development of number sense involves the ability to understand number-based situations and the effects of manipulations on numbers (Bisaillon, 2023). Research offers insights into these early mathematics skills and highlights the importance of teachers implementing effective instruction during the early years (Jacobi-Vessels et al., 2014). Manfra, Dinehart, and Sembiante (2014) found that preschool children who could count higher numbers of objects were at a greater advantage for later mathematics achievement, and those who could accurately recite numbers beyond their ability to count objects had an additional advantage.

A pilot project of Counting Collections in 2017/2018 involved delivering a two-day version of the CPD programme in 21 early years settings with 600 children aged two to five across the East Midlands (Gripton and Pawluch, 2021). In this small-scale research based on teacher interviews and case study reports, Counting Collections supported the development of teacher subject knowledge, rooted in developmental progressions in number, which led to increased progress and interest in number from the children as perceived by teachers. While the pilot study reported positive findings, the limited scope provided no opportunity to examine the transference of number skills into other learning situations or the retention of number skills over time.

Other relevant evidence includes Let's Count, a preschool mathematics intervention implemented from 2012 in disadvantaged communities across Australia. Findings indicated that learning and confidence improve when adults notice children's work in mathematics (Perry et al., 2016). The intervention involved a community approach to enhancing the mathematics learning and dispositions of young children in the year before they start school. The focus for the intervention was to work with children who lived in communities facing multiple disadvantages (Perry et al., 2016). The programme involved professional learning for early childhood educators to enhance mathematics learning and teaching and strengthen partnerships between early years educators and parents. The intervention focused on building awareness among educators and parents about how to promote children's learning through noticing, exploring, and talking about the mathematics in everyday activities (Perry et al., 2016). Recognising the moment when a child shows

interest in an activity was a key element. The intervention challenged the adult's perception of what mathematics might be and offered more opportunities for the children to engage in mathematics learning.

In another relevant study, the Numbers Count intervention was evaluated through a study of 8,000 low-achieving six- and seven-year-old children in the U.S.A. (Henry et al., 2007). After an average of 43 half-hour, one to one lessons over three months, their number age test scores had risen by 14 months with an effect size of 0.85. Pupil attitudes towards learning mathematics, measured through teacher perceptions using a bespoke survey rather than pupil self-reports because of the age of the children (Henry et al., 2007), also improved substantially, with an effect size of 0.7. Children made strong progress irrespective of their background. The success of the intervention was attributed to its design, the teacher professional development programme, and the rigorous quality assurance (Dorwick, 2015). Despite these positive findings, transference and retention of number skills were not studied.

The efficacy trial reported here consists of two strands. The impact evaluation uses pre- and post-intervention pupil number assessments conducted in schools to accurately assess pupil ability in number before and after the programme was delivered. The implementation and process evaluation (IPE) draws on data from a teacher survey and teacher interviews to gauge number confidence in pupils before and after implementation of the programme and to understand existing practice in schools. Training observations were also carried out to examine how the CPD is received by teachers.

Intervention

The intervention is described here using the TIDieR framework. The logic model diagram can be found in Appendix D.

Name

Counting Collections.

Why—theory/rationale

Counting Collections is a practical early number development programme that allows pupils to make connections between concepts through structured collaborative counting and recording sessions. The weekly sessions are based on a four-part routine, with pupils working in pairs. A key factor in the success of the intervention is the role of the adult as the facilitator and their knowledge of teaching and learning in number, including developmental progressions (learning trajectories). A small-scale pilot study (Gripton and Pawluch, 2021) indicated that Counting Collections supported the development of teacher subject knowledge, rooted in developmental progressions in number, which led to perceived increases in pupil attainment and interest in number. This warranted further investigation, especially around the varying degrees of teacher knowledge and the training needed.

Who—recipients

The intervention is designed for reception pupils (aged four to five years).

What—materials

A typical Counting Collections session follows a teacher-led sequence that guides pupils to work in pairs to plan, count, and record their approach to counting a set of manipulatives ('collections' of everyday objects such as straws, small soft toys, pinecones). These collections form a Counting Library in the classroom. There are tools to aid counting (such as boxes, jars, and number frames). The manipulatives and tools are provided by the delivery team so that all settings have the same resources available for the activities.

What—procedures

A structured routine for counting is taught in four parts: choose, strategise, count, record. This routine represents one of the core components of the programme. Collection sizes are increased over time alongside teaching of more complex strategies including counting unitised groups.

Who—provider

One teacher per school took part in the professional development programme provided by the developer. The first session was in person and was followed by four online sessions. A dedicated online learning environment known as Counting Collections Online provided ongoing support throughout the programme. Teachers could revisit content, communicate with developers, and discuss Counting Collections with other teachers.

How—format

Trained teachers delivered the intervention to pupils in class. If TAs were allocated to support these sessions, it was expected that teachers would train them so that they understand the intervention. However, TA participation in Counting Collections was optional for schools.

Where—location

The intervention was delivered in reception classrooms at schools in England. Any mainstream maintained school with at least 20 pupils in the 2023/2024 reception cohort was eligible to take part in the trial. Recruitment was restricted to three localities (North East, East Midlands and South Yorkshire, and South West). Recruitment targeted Education Investment Areas (EIAs) in line with the DfE Accelerator Fund requirement for 50% of schools to be in EIAs. Ultimately 69% of schools were recruited from EIAs.

When and how much—dosage

The intervention was delivered during two terms of the 2023/2024 school year. Teacher training took place shortly before autumn half term and the first sessions with pupils began soon after that. Delivery continued until the summer half term. Ideally, sessions were held weekly during this time. It was necessary for a minimum of 20 sessions to be completed over the study period for a school to be classed as compliant.

Tailoring—adaptation

The intervention was not tailored or adapted for the trial. The training clearly outlined the importance of the four-part routine and the requirement to maintain the delivery pattern of the learning sequence. Teachers were asked to ensure that they delivered weekly 30-minute sessions to the whole class and that children worked in pairs. The teachers were asked to set up a Counting Collections library for the resources This included a display of the counting collection boxes; the majority of the schools chose to use an open access shelved unit near a carpet area or table. It was suggested that the teachers chose how many boxes they would like to put out on display, but they were encouraged to have a variety of collection sizes and to let the children to choose their own. There was some flexibility around the pairing of the children, how the session fitted in with the business-as-usual practice, and when the session took place. The developers were responsive to teacher feedback around the practicalities of the intervention and suggested adaptations to support effective implementation. Observations of training sessions found that teachers maintained the four-part routine but shaped the delivery to fit with their classroom and practices. Details on how implementation varied between teachers and schools are discussed in the Results section.

Evaluation objectives

Impact evaluation research questions

RQ1 What is the impact of Counting Collections on reception pupil attainment in number as measured by the Sandwell Early Numeracy Test?

RQ2 What is the impact of Counting Collections on disadvantaged reception pupil attainment in number as measured by the Sandwell Early Numeracy Test?

Implementation and process evaluation research questions

IPE RQA How effectively does the training equip teachers to deliver the intervention and improve their understanding of key number concepts?

IPE RQB Is the intervention associated with improvements in teacher knowledge of learning trajectories?

IPE RQC Do teachers believe that the intervention is associated with short term improvements in pupil confidence, enjoyment, and attainment in number?

IPE RQD What level of fidelity is observed during the trial and what influences fidelity?

IPE RQE What do the trial findings indicate about scalability?

The protocol and statistical analysis plan (SAP) are both published on the EEF website.

Ethics and trial registration

Schools were approached by the developer and signed a memorandum of understanding to confirm their place on the study. This document can be found in Appendix E, along with the parent/carer information sheet explaining the trial to participants. The evaluation was approved by the SHU ethics committee on 12 January 2023 (Ethics Review ID: ER50653704). The trial is publicly registered: ISRCTN96349771.

Data protection

Sheffield Hallam University and the University of Nottingham are both data controllers for the Counting Collections evaluation. As personal data is being processed for the purpose of academic research, the main aim of which is to improve numeracy among school pupils, the legal basis is as a 'Public Task' under GDPR Article 6 (1e).

Pupil free school meals (FSM) status is not special category data under the GDPR but the DfE requires it to be treated as such. Therefore, it is processed for the purpose of research under GDPR Article 9 (j). This data was accessed from the National Pupil Database (NPD) using Unique Pupil Numbers provided by schools and was processed to determine if the Counting Collections programme has different effects on different pupil groups. The EEF was established with a remit to break the link between family background and educational attainment and all EEF projects conduct subgroup analysis on FSM pupils.

Upon completion of the EEF evaluation, SHU and UoN will retain participants' data for research and knowledge-exchange purposes, including presentations at professional or academic conferences, or publications in professional or academic journals, for up to 25 years and for a period of no less than seven years after the research project finishes. Data will be stored in the EEF data archive after publication of the final report. The project privacy notice contains details of all data to be processed.[1]

SHU's privacy notice provides full information of policies and procedures in relation to the personal data of research participants.[2]

[1] https://www.shu.ac.uk/-/media/home/research/sioe-rke/privacy-notices/eef-counting-collections-privacy-notice.docx
[2] https://www.shu.ac.uk/about-this-website/privacy-policy/privacy-notices/privacy-notice-for-research

Project team

Evaluation team—Sheffield Hallam University

Dr Martin Culliney	Senior Research Fellow at SHU: principal investigator and impact evaluation lead.
Dr Karen Daniels	Associate Professor at SHU: IPE lead.
Joanne Robson	Senior Lecturer in Primary and Early Years Education: early maths lead, IPE design, delivery, and reporting.
Sean Demack	Principal Research Fellow: statistical advisor.
Catherine Hathaway	Lecturer in Primary and Early Years Education at SHU: IPE fieldworker.
Jacqueline Launders	Senior Lecturer in Primary and Early Years Education: IPE fieldworker.
Fiona Storey Senior	Lecturer in Primary and Early Years Education: IPE fieldworker.

Development team—University of Nottingham

Dr Catherine Gripton	Associate Professor in the School of Education: principal investigator leading the University of Nottingham team.
Professor Shaaron Ainsworth	Research and Knowledge Exchange lead in the School of Education at Nottingham and Convenor for the Learning Sciences Research Institute.
Professor Geoffrey Wake	Emeritus Professor of Mathematics Education at Nottingham.
Dr Marie Joubert	Research Fellow in the Observatory for Mathematical Education at Nottingham.

Methods

Trial design

Table 3: Trial design

Trial design, including number of arms		Two-arm, two-level cluster randomised controlled trial.
Unit of randomisation		School.
Stratification variable(s) (if applicable)		Geographical area (North East, South West, East Midlands/South Yorkshire); existing use of Mastering Number intervention.
Primary outcome	Variable	Number attainment.
	Measure (instrument, scale, source)	GL Sandwell Early Numeracy Test (B) score (0–75).
Secondary outcome(s)	Variable(s)	N/A*
	Measure(s) (instrument, scale, source)	N/A*
Baseline for primary outcome	Variable	Number attainment.
	Measure (instrument, scale, source)	GL Sandwell Early Numeracy Test (A) score (0–75).
Baseline for secondary outcome(s)	Variable	N/A*
	Measure (instrument, scale, source)	N/A*

* No secondary outcomes used in this evaluation, see Outcome Measures, page 12.

This was a two-arm, two-level clustered efficacy trial with pupils clustered into schools. The unit of randomisation was schools. Intervention schools delivered Counting Collections to pupils in the participating reception class during the study period and were offered a £200 payment for taking part, while control schools continued with business as usual and were offered £500. Both groups were paid in two instalments: after completion of baseline and outcome testing. There was no waitlist. Baseline testing was carried out from September to November 2023 and outcome testing in June and July 2024. Both involved assessors working for the evaluator visiting schools to conduct the Sandwell Early Numeracy Test (SENT-R) with pupils. At baseline, SENT-R version A was used and at outcome it was SENT-R version B. Further details on pupil testing can be found in Outcome Measures (page 12).

Participant selection

Schools were recruited by the developer. Participation was conditional upon abiding by the terms of the MoU. All mainstream schools with at least 20 pupils in the 2023/2024 reception cohort were eligible for the trial. The aim was to recruit at least 50% of participating schools from EIAs. This target was met (Table 4). Schools were allowed to choose one reception class to take part. It would have been burdensome for schools to source cover for more teachers to attend

the training and substantial class-level effects were not expected with pupils of this age owing to the relative rarity of ability grouping for early primary school years (Demack, 2019).

Counting Collections is a whole-class intervention and there were no constraints on pupil eligibility other than the requirement to be in a reception class (age four to five) throughout the study year. As assessing more than 20 pupils per school would have increased costs and burden on schools while bringing only minor improvements in statistical sensitivity it was agreed to limit baseline assessments to 20 pupils per school. For larger classes, 20 pupils were randomly selected by the evaluators prior to baseline testing. Those pupils formed the study sample. There was no reserve list or procedure to replace pupils withdrawing from the trial.

Outcome measures

Baseline measures

SENT-R, a one to one numeracy test for pupils aged four to eight years, was used as the baseline and primary outcome measure. It has two components (A and B) and is suitable for gauging the impact of classroom interventions on a pre- and post-test basis. SENT-R A is used as the baseline measure; SENT-R B the outcome measure. SENT-R is published by GL Assessment, which confirmed that it is appropriate to use it in this way. It has been used similarly in previous research (Torgerson et al., 2011:49).

The assessment explores five strands of basic numeracy skills: identification, oral counting, value, object counting, and language, but these are not validated for use as standalone measures and were therefore not considered suitable as secondary outcomes. As Counting Collections aims to improve number attainment overall, the SENT-R assessment is relevant to the intervention. Baseline testing was conducted in schools—by researchers blinded to group allocation—between 25 September and 24 November 2023.The SENT-R assessment provides an opportunity to carry out analysis by age-equivalent band and is administered on a one to one basis using scripted questions, images, and manipulatives. The assessment stops when the pupil has made two consecutive errors in all of the strands. The researchers were students on education and psychology degrees recruited by the subcontractor appointed by the evaluators to administer pupil assessments. To cover a shortage in one of the study regions, toward the end of the assessment period ten teaching assistants were hired through a supply agency to work as assessors. They undertook the same training as the others. This involved an introduction to the study, principles for working with children in schools, data protection, and how to administer the assessment, including a practical demonstration followed by paired role play.

Primary outcome

As described, SENT-R B was used as the primary outcome measure. Data collection took place between 3 June and 22 July 2024. It was led by the evaluators in the same way as the baseline testing.

Secondary outcomes

No secondary outcomes are used in this evaluation. The separate scales from the SENT-R assessment are not validated for standalone use and the intervention does not aim to prioritise one strand of number learning over others. It was therefore agreed during the setup period that no secondary outcome would be used.

Sample size

The design was a two-level, clustered RCT. In calculating the Minimum Detectable Effect Size (MDES)—the smallest effect size that could be detected as statistically significant at $p < 0.05$ with a statistical power of 80% or higher—our estimates were based on the following assumptions:

- **$M_{j\text{-}k\text{-}2}$** T-distribution multiplier assuming a two-tailed test with a statistical significance of 0.05, statistical power of 0.80 and J-K-2 (175) degrees of freedom;
- **R_i** participant (pupil) level pre/post-test correlation of 0.6 ($R_i^2 = 0.36$);
- **R_c** cluster (school) level pre/post-test correlation of 0.2 ($R_c^2 = 0.04$);

ρ intracluster correlation (ICC) 0.17;

j number of schools (180);

m number of pupils per school (20);

k number of cluster-level covariates (3); and

P proportion of schools allocated to intervention group (P = 0.5).

The ICC and participant correlation values were taken from the evaluation of Maths Champions (Robinson-Smith et al., 2018), the only early years mathematics trial published by the EEF at the time of writing the protocol. The ICC reported at the analysis stage of that trial was 0.17, and the pupil pre/post-test correlation was 0.59. These figures come from an evaluation using a different outcome measure with slightly younger children than those participating here, but nevertheless this is recent evidence from a trial of the same duration and on the same subject. Cluster-level correlations were not supplied by Robinson-Smith et al. but were conservatively estimated in the protocol for this trial as 0.20. Calculations were performed in Excel using the formula set out in Bloom et al. (2007) for two-level clustered randomised controlled trials. This allows covariates to be included at both individual (pupil) and cluster (school) level, which in turn increases sensitivity.

Equation 1: Minimum Detectable Effect Size in a two-level clustered randomised controlled trial

$$\mathrm{MDES} = M_{j-k-2}\sqrt{\left(\frac{\rho(1-R_c^2)}{\mathrm{P}(1-\mathrm{P})\mathrm{J}}\right)+\left(\frac{(1-\rho)(1-R_i^2)}{\mathrm{P}(1-\mathrm{P})\mathrm{Jm}}\right)}$$

Based on the assumptions above, an MDES of 0.20 standard deviations would require 150 schools with 20 pupils each. However, to allow for potential attrition, 180 schools were recruited. This was the sample size at randomisation, when the overall MDES was calculated as 0.18.

The subgroup of FSM pupils, defined by the EVERFSM_6_P NPD variable as per EEF guidance, was estimated at eight per school. The number was expected to be higher than average in Education Investment Areas, where the trial aimed to recruit at least 50% of participating schools, and ultimately 69% of recruited schools were in EIAs (n = 125, 62 intervention and 63 control). At protocol, with 180 schools, the FSM MDES was 0.20; with 150 schools it was 0.22. At the analysis stage, there were 650 FSM pupils, with an average of 3.8 (rounded up to four) per school, which was lower than expected although above national average for the reception year group.[3] These figures are presented in Table 7 (Impact Evaluation, p. 21).

Randomisation

Randomisation was carried out at school level to minimise spillover risk. The `stratarand` command in Stata 17 was used. Geographical region—East Midlands and South Yorkshire, North East, and South West—was used as a stratification variable to reduce the risk of allocation imbalance in certain geographical areas undermining the viability of the training. The other stratifier was whether the school was already using Mastering Number, a government-funded national programme used widely in schools in England. Excluding schools that are using Mastering Number would have severely restricted recruitment and imbalance between the intervention and treatment groups with regard to use of Mastering Number could have confounded the results of this trial. The evaluators were blind to group allocation at the time of randomisation but there was no intention to maintain this after schools were allocated as the evaluators needed to inform the development team of which schools were in the intervention group so they could receive the programme resources and be invited to the training.

[3] https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics#dataBlock-892e8acf-47ca-4abc-b337-38a0bbcf9e6c-tables

Randomisation was completed on 14 July 2023. Parent information sheets were sent to schools on 5 September. Schools began submitting pupil data to the evaluators on 18 September. Baseline pupil assessments began on 25 September. Teacher training for intervention schools began on 9 October, while pupil data collection and baseline assessments were still underway. Schools were already aware of their allocation by the time any pupil data was collected or baseline assessments completed. This was deemed unavoidable given the lack of time to complete all evaluation and delivery activities, including the postage of Counting Collections materials to intervention schools.

Statistical analysis

Primary analysis

Multilevel linear regression models are estimated for the SENT-R B primary outcome, with pupils clustered into schools. For each model, the coefficient of the treatment allocation variable, which distinguishes between intervention and control group pupils, is converted into Hedges' g effect size statistics with 95% confidence intervals. The first model only includes the treatment allocation identifier (an outcome only model). The second model adds the baseline test score as a covariate at the pupil and school levels.[4] SENT-R A is used as the baseline covariate for analysis of the primary outcome. The final model also includes the stratifiers used in the randomisation process (geographical area, whether the school is using the Mastering Number intervention) and provides the headline intention to treat (ITT) impact analysis for the SENT-R B primary outcome. The headline effect size is calculated using the coefficient from the following equation:

Equation 2: ITT analysis model

$$Y_{ij} = b_0 + b_1 Allocation + b_2 Baseline_{ij} + b_3 Baseline_{school\,j} + b_4 Mastering\ Number_j + b_5 North\ East_j + b_6 South\ West_j + u_j + e_{ij}$$

Where Y_{ij} is the outcome for pupil *i* in school *j*, b_0 is the constant, and *Allocation* is a binary indicator of school treatment allocation. Pupil and school level baseline covariates are represented by *Baseline* and *Baseline_school*.

The stratifiers used in the randomisation are denoted as *Mastering Number,* a binary indicator of whether the school is using that intervention, while *North East* and *South West* are dummies derived from the region variable. The random intercepts are represented by u_j and e_{ij} is the error term.

Follow-on ITT analyses examine the impact of Counting Collections on number attainment among disadvantaged pupils, as defined by the NPD variable EVERFSM_6_P (here, this simply refers to FSM eligibility in the study year as the pupils are in the reception phase). The same approach used for the headline ITT analyses is followed.

Secondary analysis

No secondary outcomes were used in this evaluation.

Analysis in the presence of non-compliance

Compliance is measured at the school level. Full details of the three indicators are provided below. Each relates to activities undertaken by the participating teacher, yet as only one teacher per school took part in the trial, these are effectively school-level indicators. The measures are combined into overall full and part compliance indicators at the school level and then used to estimate the Complier Average Causal Effect (CACE). The purpose of CACE is to estimate the impact of Counting Collections for pupils deemed to have 'complied' with the intervention, though as no pupil-level compliance indicators are being used, pupil compliance is simply being at a school classed as compliant.

The developer specified three criteria for schools to meet to achieve full compliance:

[4] These will be centred so that the school level will be centred on the mean for all schools and the pupil level will be centred around the school mean.

- teacher attendance at all five professional development sessions (first session in person, then synchronous online sessions);
- graded manipulatives and supporting tools must be present in the classroom (teacher self-report); and
- teaching at least five weekly Counting Collections sessions per half term (20 in total).

If one of the online professional development sessions was accessed asynchronously, full compliance is still achieved provided that the other criteria are met. Part compliance can be achieved through teachers completing the online CPD sessions asynchronously. The unit of analysis for compliance is the school, although, as mentioned above, only one teacher per school took part in the trial.

CACE is estimated using two-stage least squares (2SLS) regression. The first stage models the compliance variables using the same explanatory variables used for the headline ITT analyses along with additional school-level items that are available via the school census as included in Table 3. This is a multilevel logistic regression model used to generate predicted compliance for use in the second stage model. The second stage models predicted compliance in place of the group identifier variable in the ITT analyses specified above to generate the CACE estimates. The intention was to undertake this process twice, for full and part compliance, although ultimately this was not necessary as no schools were classed as partly compliant (see Results section, below).

Missing data analysis

The impact analyses examine missing data in the outcome and explanatory variables and consider whether it is reasonable to assume that the missing data is missing at random. A multilevel logistic regression model with a binary outcome denoting when outcome data is missing (‘1’) or not (‘0’) and the same covariates as the headline ITT model is estimated to identify any patterns in missingness. This model is then replicated with only pupils at schools that took part in the outcome testing to focus on pupil-level attrition in those schools.

The SAP stated that in the instance of any missing outcome data, the baseline and ITT samples would be compared across all ITT variables and additional variables shown in Table 9. With over 5% of pupil outcome data missing, a multilevel logistic regression model with a binary outcome identifying when outcome data is missing (‘1’) or not (‘0’) would be estimated, and the ITT variables and additional school-level variables are used to identify whether the missing outcome data can be assumed to be missing completely at random. If none of the explanatory variables had been found to account for a statistically significant amount of variation in the missing data outcome, it would have been cautiously assumed that the data is missing completely at random, but as this was not the case multiple imputation was used and the results compared with the headline ITT analysis for the primary outcome.

If one or more explanatory variables are found to account for a statistically significant amount of variation in the missing data outcome, we would undertake a sensitivity analysis to repeat the ITT analysis with these variables included. The potential bias introduced by missing outcome data on the ITT estimate will be illustrated by comparing the estimated ITT effect size with the effect size estimated from the ITT model including the additional variables.

Sub-group analyses

Pupils eligible for free school meals as identified by the EVERFSM_6_P NPD variable are analysed below as a separate subgroup, as is required for all EEF trials. This data is taken from the NPD for consistency with other EEF trials. The second research question for the trial is based on this subgroup analysis, although it is exploratory as the study is not powered for subgroup analysis. Further analysis of the whole study sample, with an interaction effect between FSM status and treatment allocation included in the model, is also presented below.

Estimation of effect sizes

The effect size measure used was Hedges’ g. This is calculated using the following equation.

Equation 3: Effect size calculation

$$ES = \frac{(T - C)_{adjusted}}{\sqrt{\delta^2_{sch} + \delta^2_{pup}}}$$

Where δ^2_{sch} is the school-level variance and δ^2_{pup} is the pupil-level variance for the language outcome from the empty/null multilevel model. $(\text{T} - \text{C})_{\text{adjusted}}$ is the mean difference between the attainment of pupils in treatment schools and pupils in control schools. This is obtained from the coefficient for the treatment allocation variable from the final analyses.

Estimation of ICC

For the primary outcome at both pre- and post-intervention, ICCs at the school level were estimated using a null (empty) two-level multilevel variance components model. Within the analyses, a table presents the variance decomposition for the two levels (school and pupil) along with the ICC estimates.

Equation 4: ICC calculation

$$\text{ICC} = \frac{\text{Variance}_{\text{school}}}{\text{Variance}_{\text{school}} + \text{Variance}_{\text{pupil}}}$$

Implementation and process evaluation[5]

Research methods

The IPE design was based on the evidence-informed logic model, agreed by the delivery team and evaluation team. The significance of teacher knowledge of number learning and how this may affect pupil knowledge, confidence, and enjoyment of number was explored. IPE activity comprised:

1. Observation of six training sessions

Six training sessions were observed by early mathematics specialists to examine the content and format of the training and materials. This aimed to improve understanding of the programme and inform the school-based fieldwork and post-intervention surveys. Records of training attendance were also analysed to gauge fidelity.

2. Pre- and post-intervention online teacher survey

Pre- and post-intervention online teacher surveys explored 'business as usual', school context, mathematics training, teacher experience, and confidence. The post-intervention survey was sent to all schools and included additional questions for the intervention group. The baseline assessment period was demanding for schools as it included staff training in intervention schools along with the requirement to submit pupil data and undertake the assessments. The evaluation team did not think it was appropriate to run the pre-intervention teacher survey during this period. Instead of delaying this until after baseline data collection, it was agreed with the delivery team that sending the pre-intervention survey after the training had begun and the programme was being implemented in schools would produce misleading data that would be likely to understate any changes in teacher responses associated with the intervention. As such, the pre-intervention survey was cancelled and the post-intervention survey was designed to address the relevant research questions. The questionnaire can be found in the further appendices to this report.

[5] See IPE guidance for further details.

3. School visits

School visits were undertaken to assess fidelity and ascertain influences on implementation and the extent to which the teacher practice is aligned with expectations of the programme. Our team of early mathematics experts visited ten schools across the regions where Counting Collections was implemented. Fieldwork involved:

- observation of at least one teacher-led, whole-class Counting Collections session;
- observation of at least one pupil play number interaction;
- examination of the resources available in the 'maths area' of the classroom; and
- an interview with the teacher and the school leader responsible for Counting Collections at school level.

The interview and observation schedules can be found in the further appendices to this report.

Analysis

Table 5 (below) illustrates how the data collection methods map onto the research questions. Descriptive statistics based on survey responses are presented to gauge participant perceptions of confidence in mathematics teaching and, for respondents in intervention schools, experiences of the training and of delivering the programme. The Chen et al. (2014) questionnaire on early years mathematics teacher beliefs and confidence is adapted for the post-intervention survey. Pertinent themes arising from the fieldwork visits and interviews with teachers and school leaders were also included in the post-intervention survey where relevant.

Fieldwork interviews were one to one and semi-structured. An interview schedule guided by the main research questions and pre-intervention survey findings was designed. Interviews enabled the evaluation team to ask how Counting Collections is working, for whom, and under what circumstances. Teacher perceptions on the impact of Counting Collections on pupil progress in number and any attitudinal changes were also explored, along with any barriers to its successful implementation. Interview data was coded using NVivo. A variable-orientated approach to case analysis was taken to identify broad patterns and recurring themes across cases (Miles et al., 2019).

An observation schedule was prepared for lesson observations. This examined fidelity to the Counting Collections session routine, content, and key pedagogical principles. Descriptive fieldnotes were taken to describe the ways in which a mathematics area is organised and the materials it contains for pupils. Analysis focused on identifying key features of manipulative and tool provision in classrooms and examples of good practice.

Table 5: IPE methods overview

Research question	Data collection method	Participants
How effectively does the training equip teachers to deliver the intervention and improve their understanding of key number concepts?	Training observations Interviews and observations in school Post-intervention survey	Teachers, trainers Teachers in 10 intervention schools Teachers in all intervention schools
Is the intervention associated with improvements in teacher knowledge of learning trajectories?	Interviews and observations in school Post-intervention survey	Teachers in 10 intervention schools Teachers in all intervention schools
Do teachers believe that the intervention is associated with short term improvements in pupil confidence, enjoyment and attainment in number?	Interviews and observations in school Post-intervention survey	Teachers in 10 intervention schools

		Teachers in all intervention schools
What level of fidelity is observed during the trial and what influences fidelity?	Interviews and observations in school Post-intervention survey	Teachers in 10 intervention schools All intervention teachers
What do the trial findings indicate about scalability?	Interviews and observations in school Post intervention online survey	All intervention teachers
How effectively does the training equip teachers to deliver the intervention and improve their understanding of key number concepts?	Training observations Interviews and observations in school Post-intervention survey	Teachers, trainers Teachers in 10 intervention schools Teachers in all intervention schools

Costs

Cost information was provided by the development team to reflect what it would cost a school looking to implement Counting Collections outside of this trial. Costs per pupil per year were calculated over a three-year period based on EEF guidance. All figures are presented in the Cost section (page 47).

Timeline

Table 6: Timeline

Dates	Activity	Staff responsible/leading
Oct–Nov 22	Set-up meetings and IDEA workshop	All
Nov–Dec 22	Ethical approval Draft MoU, consent and information forms Design IPE instruments Evidence review	SHU
Feb 23	Protocol Trial registration	SHU
Feb–Jun 23	Recruitment Data collection from schools	UoN
Jul 23	Randomisation	SHU
Sep–Nov 23	Collect pupil data Baseline testing	SHU
Nov 23	CPD session 1 teacher training/observations	SHU/UoN/schools
Nov 23–May 24	Intervention delivery	Schools
Nov 23–May 24	Conduct IPE school visits	SHU/schools
Nov 23–May 24	CPD sessions 2–5 teacher training/observations	SHU/UoN/schools
Jan 24	SAP	SHU
Mar 24	NPD application	SHU
Jun–Jul 24	Outcome testing Post-intervention teacher survey	SHU/schools
Sep 24–Jan 25	Data analysis	SHU

Impact evaluation

Participant flow including losses and exclusions

Figure 1: Participant flow diagram

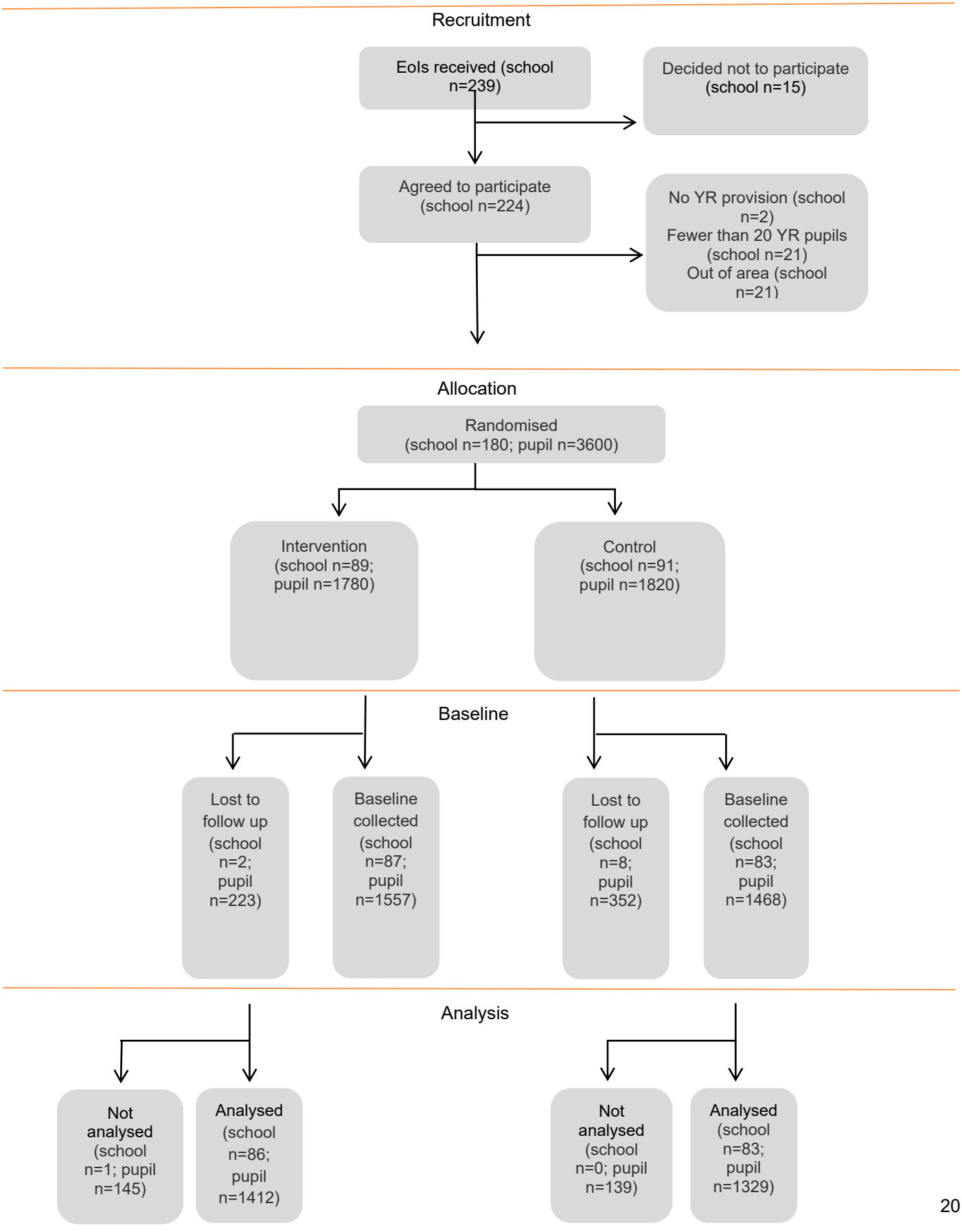


Table 7: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.18	0.20	0.18	0.20	0.16	0.20
Pre-test/post-test correlations	Level 1 (pupil)	0.60	0.60	0.60	0.60	0.73	0.71
	Level 2 (class)	-	-	-	-	-	-
	Level 3 (school)	0.20	0.20	0.20	0.20	0.56	0.67
Intracluster correlations (ICCs)	Level 2 (class)	-	-	-	-	-	-
	Level 3 (school)	0.17	0.17	0.17	0.17	0.16	0.21
Alpha *		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		2	2	2	2	2	2
Average cluster size		20	8	20	8	16	4
Number of schools **	Intervention	90	90	89	89	86	75
	Control	90	90	91	91	83	77
	Total:	180	180	180	180	169	152
Number of pupils ***	Intervention	1800	720	1780	712	1412	326
	Control	1800	1800	1820	728	1329	324
	Total:	3600	1440	3600	1440	2741	650

* Please adjust as necessary for trials with multiple primary outcomes, three-arm trials, etc., when a Bonferroni correction is used to account for family-wise errors.
** Adjust as necessary for trials that randomise at the class or other level.
*** Please adjust as necessary, for example, for trials that are randomised at the class level.

Attrition

The trial recruited 180 schools. As discussed elsewhere in this report, randomisation was conducted prior to baseline testing and the collection of pupil data from schools. Attrition is normally calculated as the ratio between the number of pupils randomised and the number analysed. In this case it was necessary to estimate the number of pupils randomised: as randomisation took place before pupil data collection and baseline testing, the exact number of pupils was not known at randomisation. It was estimated that 3,600 pupils were in the study sample as a limit of 20 pupils per school was imposed. When using this estimated figure, the attrition rate is 23.9% (see Table 8).

It is worth reiterating that eight schools did not respond to requests for the required pupil data and were withdrawn from the study between randomisation and baseline testing. Six of these were from the control group, explaining some of the imbalance in attrition. The remaining 172 schools sent the required data on 3,340 pupils. If the attrition was calculated using this figure it would be 18%. A total of 3,025 pupils completed baseline testing and 9.4% of these did not take part in outcome testing. These details are presented above in Figure 1 along with a breakdown according to treatment allocation.

Table 8: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	1780	1820	3600
	Analysed	1412	1329	2741
Pupil attrition (from randomisation to analysis)	Number	368	491	859
	Percentage	20.7%	27.0%	23.9%

Pupil and school characteristics

Table 9 shows the characteristics of the schools randomised at the start of the trial. Schools were recruited in three regions: the North East, the South West, and the East Midlands (later extended into South Yorkshire). The East Midlands/South Yorkshire group accounted for over half (53%) of participating schools, with the remainder divided almost evenly between the other two regions. Just over half of schools (53%) reported using the Mastering Number intervention. Both of these variables were well balanced between the intervention and control groups having been used as stratifiers in the randomisation.

School OFSTED rating was not used as a stratifier, yet the sample is balanced on each performance category, with eight outstanding schools in both the intervention and control groups and very similar numbers of good schools and those requiring improvement represented. The total number of pupils in each school was almost equal in intervention (340) and control (351) schools, with both being larger than the national average of 288 pupils. Schools in the study sample also had a higher percentage of disadvantaged pupils (29%) than the national average (23%) although the intervention and control groups were well balanced on this.

Pupil baseline assessment scores were slightly higher in control schools (mean = 17.69, SD = 8.67) than intervention (mean = 16.97, SD = 8.21), with an effect size of -0.09. Further details are presented along with the impact evaluation results. Histograms showing the distribution of the raw baseline and outcome scores can be found in Appendix B.

Table 9: Baseline characteristics of treatment groups as randomised

School level (categorical)	**National level mean**	**Intervention group**		**Control group**		
		n/N (missing)	**Count (%)**	**n/N (missing)**	**Count (%)**	
School level (categorical)						
East Midlands/South Yorkshire	N/A	48/89	48 (53.9%)	48/91	48 (52.7%)	
North East	N/A	21/89	21 (23.6%)	22/91	22 (24.2%)	
South West	N/A	20/89	20 (22.5%)	21/91	21 (23.1%)	
Uses Mastering Number	N/A	47/89	47 (52.8%)	48/91	48 (52.7%)	
Does not use Mastering Number	N/A	42/89	42 (47.2%)	43/91	43 (47.3%)	
OFSTED ratings						
Outstanding	12.1%	8/89	8 (9.0%)	8/91	8 (8.8%)	
Good	79.3%	65/89	65 (73.0%)	64/91	64 (70.3%)	
Requires improvement	7.9%	4/89	4 (4.5%)	5/91	5 (5.5%)	
Missing	N/A	12/89	12 (13.5%)	14/91	14 (15.4%)	
School level (continuous)		**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	
Total number of pupils (including part-time pupils)	288(194)	86/89(3)	340 (188)	89/91 (2)	351 (243)	
Percentage of disadvantaged pupils	22.8(15.11)	86/89(3)	29.37 (16.34)	89/91 (2)	29.69 (18.438)	
Pupil-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size
SENT-R Baseline		1557	16.97 (8.21)	1468	17.69 (8.67)	-0.09

Outcomes and analysis

Primary analysis

The primary outcome for this trial was SENT-R B and the baseline measure was SENT-R A. There were no secondary outcomes to be measured. At baseline, pupils in control schools achieved higher baseline scores (mean = 17.69, SD = 8.67) than counterparts in intervention schools (mean = 16.97, SD = 8.21, see Table 9), however, outcome test scores among intervention pupils were marginally higher (26.48) than for control pupils (26.24). Histograms showing the distribution of the assessment scores are presented in Appendix C along with further details on effect size calculations.

Table 10 displays the results from the headline analysis undertaken with the ITT sample. There is a positive effect of 0.08 associated with the intervention, equivalent to one month of additional progress. The lower confidence interval is slightly below zero (lower CI = -0.015, upper CI = 0.175) and the p-value (0.097) is above 0.05. As the trial was not powered to detect an effect of this magnitude there is a degree of uncertainty over this finding. At protocol and randomisation, the power calculations showed that the ITT sample was powered to detect an effect size of 0.20. The assumptions proved to be conservative as repeating the power calculations with the analysis sample produced a lower MDES of 0.16, the difference largely due to the pupil and school pre-post correlations on the outcome measure being higher than anticipated (see Table 7 above). However, to detect the observed effect size (0.08) at the $p < 0.05$ level would have required a sample of 586 schools if all other parameters remained the same.

Table 10: Primary outcome

	Unadjusted means				**Effect size**		
	Intervention group		**Control group**				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
SENT-R Outcome	1412(308)	26.48 (25.971; 26.981)	1329 (291)	26.24 (25.729; 26.76)	2741 (1412; 1329)	0.08 (-0.015; 0.175)	0.097

Analysis in the presence of non-compliance

Of the 89 schools randomly allocated to the intervention group, two withdrew from the study before baseline testing and did not take part in any of the Counting Collections teacher training or deliver the programme to their pupils at all. Compliance with the intervention was assessed on three criteria. First, schools were required to ensure that the intervention materials were present in the classroom. It was confirmed by the developer that all intervention schools had evidenced this during the training sessions. Schools were also required to deliver at least one Counting Collections session per week over at least 20 weeks. Again, this was achieved by all 87 intervention schools.

The CPD programme comprised five training sessions for teachers. For the school to be rated as compliant, it was necessary for the participating teacher to attend each of these sessions. It was agreed that teachers attending all sessions but attending more than one asynchronously would be classed as partly compliant. However, as the one teacher who completed any training asynchronously did only a single session in this format, they achieved full compliance. As such, no schools were classed as partly compliant with the intervention.

Compliance was high overall (79 of the 91 intervention schools randomised) but distributed unevenly between the three geographical regions. All schools in the South West region achieved full compliance, as did all but one in the North East. The other schools that were not deemed compliant to the trial were in the East Midlands/South Yorkshire region, which accounted for over half of the schools in the study (96 of 180). In terms of the other school-level variables examined above in Table 9, it is also worth noting that all schools that were not deemed compliant were rated by OFSTED as good; none were rated outstanding or requiring improvement. Again, this is likely to be due to most schools having a good OFSTED rating as opposed to any systematic variation with regard to compliance. Four of the eight non-compliant schools were delivering Mastering Number. These figures are not tabulated here to avoid identifying any schools.

Table 11: Compliance with the intervention

Outcome	N	%
Materials present in classroom	87	100
Delivery of at least 20 weekly Counting Collections sessions	87	100
Attendance at five CPD sessions	79	91
Fully compliant	79	91

To examine the factors predicting compliance using the variables presented in Table 9, a series of multilevel logit models were estimated. The first model included geographical region, use of Mastering Number, OFSTED rating, number of pupils in school, and percentage of FSM pupils as predictors. The second model omitted OFSTED rating, as all non-compliant cases were in the same category, and region. The third model was restricted to individuals completing the outcome assessment. For the second and third models, all odds ratios were between 0.9 and 1.1. None of these models succeeded in generating a predicted compliance variable. This is perhaps due to the remaining school variables performing poorly at predicting compliance, which was high overall (91%, see Table 11).

While descriptive statistics showed that pupils in compliant schools scored slightly higher on the outcome assessment (Table 12), due to the high compliance observed it is assumed that compliance had little effect on the headline ITT results. This is confirmed through per protocol analysis, which showed an effect size of 0.09 for pupils in fully compliant schools (compared to 0.08 for the headline ITT model), again equivalent to one month of additional pupil progress. Calculating the effect size using CACE (see Equation 5, based on Bloom 1984:232), produces an estimate of 0.09 standard deviations. The similarity of the effect sizes from the ITT, per protocol, and CACE analyses suggest that while compliance is associated with higher scores on the primary outcome measure, compliance was high enough that the headline result remains robust.

Table 12: Baseline and outcome test scores for pupils in compliant and non-compliant intervention schools

Compliance	Baseline	Outcome
Compliant mean	16.91	26.53
SD	8.17	9.74
N	1407	1294
Non-compliant mean	17.61	25.89
SD	8.80	9.01
N	135	118

Equation 5: CACE calculation

$$CACE\ estimate = \frac{ITT\ estimate}{proportion\ of\ pupils\ in\ compliant\ schools} = \frac{0.08}{0.916} = 0.087$$

Table 13: Per protocol analysis for primary outcome (compliant cases vs control)

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges g (95% CI)**	**p-value**
SENT-R B	1294 (241)	26.53 (25.998; 27.06)	1329 (291)	26.24 (25.729; 26.76)	2623 (1294; 1329)	0.09 (-0.007; 0.186)	0.069

Missing data analysis

This trial suffered from missing data as eight of the schools entered into the randomisation ultimately did not submit pupil data to the evaluation team or take part in the baseline testing. The exact number of pupils in the randomisation sample therefore remains unknown, and the figure used to calculate attrition is an estimate based on 20 pupils per school. Baseline data was available for 3,025 pupils in 170 schools, and 2,741 pupils completed outcome assessments. The missing data analysis is based on the pupils for whom baseline data is available as it is important to determine whether pupil number ability, as measured by performance on the baseline test, affected the likelihood of missing outcome data. This was examined through a logit model with missing outcome data as the binary outcome variable (1=missing, 0=not missing), and the same covariates as in the headline ITT model. The results are displayed in Table 1 and show that the odds ratio for the baseline measure (0.949) is below one, meaning that lower baseline scores are associated with a greater chance of missing outcome data. For the other covariates in the model, there are no other statistically significant relationships, even when additional school level covariates are included. However, as much of the attrition occurred between randomisation and baseline testing, this approach clearly has limitations.

Table 14: Logit model with missing outcome data as outcome variable

	Model 1	Model 2
Allocation	0.938(0.167)	1.020(0.187)
Raw baseline score	0.949***(0.008)	0.944***(0.009)
North East (reference East Midlands)	0.895(0.198)	0.878(0.201)
South West (reference East Midlands)	0.909(0.204)	0.816(0.200)
Mastering Number (1=Yes, 0 = No)	1.272(0.230)	1.250(0.231)
Number of pupils in school		1.000(-0.001)
Percentage FSM pupils in school		1.003(0.006)
Ofsted Outstanding (reference Good)		0.876(0.257)
Ofsted Requires improvement (reference Good)		1.133(0.424)
Constant	0.144***(0.049)	0.131***(0.058)
N pupils	3,025	2,640
N schools	170	149
Log Likelihood	-894.2	-771.4
Df	5	9
Chi2	37.20	44.01

Standard errors in parentheses. *** p < 0.001, ** p < 0.01, * p < 0.05.

Having established that lower baseline scores are associated with greater likelihood of missing outcome data, multiple imputation was used to investigate whether this may have biased the results in the headline ITT analysis. Outcome scores were imputed for the 284 cases where data on the baseline assessment and the other covariates from the headline ITT analysis were available and outcome data was missing. This was repeated ten times and the main analysis was conducted again using each of these imputed outcomes as the outcome variable. The resulting effect sizes ranged from 0.07 to 0.09, very similar to that from the main ITT model (0.08), and still consistent with one month of additional pupil progress. This indicates that attrition from baseline to outcome testing did not meaningfully affect the results of this trial.

Sub-group analyses

The headline analysis was repeated on the subgroup of 650 pupils defined as disadvantaged by the EVERFSM_6_P indicator in the NPD (Spring Census 2024). The full set of covariates shown in Equation 2 and Table 14 were included. Table 15 shows that the effect size (0.03) is noticeably smaller than for the ITT sample (0.08), and the point estimate is subject to greater uncertainty with wider confidence intervals and a higher p value (CIs: -0.118; 0.174; p value 0.709), although the smaller sample size is partly responsible. To further investigate the effect of pupil disadvantage, the ITT analysis model was repeated with the full sample, adding FSM status and an FSM*Allocation interaction term as covariates. The results are also shown in Table 15. Again, the effect size (0.02) is smaller than the headline effect size (0.08) and the trial was not powered to detect an effect of this size. While the intervention is associated with one month of additional progress in pupils, there is no evidence of additional progress for disadvantaged pupils.

To reiterate, the number of FSM-eligible pupils was not known at protocol or randomisation as the NPD data was not accessed until both baseline and outcome assessment data was available. The number of FSM pupils per school (four) was lower than anticipated (eight), and of the 169 schools in the analysis sample, only 152 had any FSM pupils although some had as many as 16. Even accounting for pupil attrition from randomisation to analysis, the assumed percentage of FSM pupils (40%) was higher than observed (25%). It is also worth noting that FSM pupils achieved lower baseline scores than the overall sample. All models presented in this section include pupil and school baseline scores as covariates. This is important to acknowledge when considering the very small difference in test scores between disadvantaged intervention and control pupils on the primary outcome, and accounts for the discrepancy between the small positive point estimate observed in contrast to the small negative difference in unadjusted means.

Table 15: FSM subgroup analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
SENT-R B: FSM subgroup	326 (1394)	22.38 (21.485; 23.27)	324 (1296)	22.93 (21.952; 23.9)	650 (326; 324)	0.03 (-0.118; 0.174)	0.709
SENT-R B: analysis sample, with FSM*Allocation interaction	1412 (308)	26.48 (25.971; 26.981)	1328 (292)	26.24 (25.729; 26.761)	2740 (1412; 1328)	0.02 (-0.19; 0.043)	0.217

Additional analyses and robustness checks

Table 16 shows the full results from the multilevel models on the ITT sample using the primary outcome. Model 1 shows that pupils in intervention schools obtained slightly higher marks on the outcome test than counterparts in control schools when controlling for no other covariates (B = 0.19). However, as the mean baseline test score was higher in control schools (see Table 9), when school and pupil baseline covariates are added in Model 2, the allocation coefficient increases to 0.79. Model 3, which also includes the stratification variables, produces a very similar allocation coefficient (B = 0.77). Pupils in the North East region achieved lower scores on the outcome test than those in the East Midlands

(B = -1.28), which was selected as the reference category as it contained the largest number of participating schools. On the other hand, pupils in the South West achieved higher scores (B = 1.71). However, these categories were evenly distributed between the intervention and control groups due to their inclusion as stratifiers in the randomisation, hence the negligible change in the allocation coefficient between Models 2 and 3.

Table 16: Results from multilevel models of primary outcome with ITT sample

	Model 1	Model 2	Model 3
Allocation	0.19 (0.602)	0.79 (0.495)	0.77 (0.467)
Centred school baseline score		0.65 (0.072)***	0.65 (0.068)***
Centred baseline score		0.87 (0.015)***	0.871 (0.015)***
North East (reference East Midlands /South Yorkshire)			-1.28 (0.577)*
South West (reference East Midlands /South Yorkshire)			1.71 (0.583)**
Mastering Number (1 = yes, 0 = no)			0.52 (0.472)
Constant	26.27 (0.431)***	25.71 (0.353)***	24.82 (0.798)***
N pupils	2,745	2,741	2,741
N schools	169	169	169
Hedges's g	0.02	0.08	0.08

Standard errors in parentheses.
*** p < 0.001; ** p < 0.01; * p < 0.05.
Hedges' g for all models calculated using unconditional variance (from Model 1, see Table 17).

Estimation of effect sizes

The effect size (0.08) was calculated by dividing the allocation coefficient, which is the mean difference between the intervention and control group outcome scores controlling for all covariates (0.77), by the unconditional standard deviation (9.651), derived from the total variance from the outcome only model. The formula was presented above (Equation 3).

Estimation of ICC

The ICC value for the outcome-only model that provided the school- and pupil-level variance was 0.108. For the full model including all covariates, used to generate the coefficient for the allocation variable that provides the numerator for the ICC equation, the ICC was 0.164 (at randomisation, 0.17 was assumed). The formula was presented above (Equation 2).

Table 17: ICC values

Model	School variance	Pupil variance	Total variance	ICC
Outcome only	10.072	83.077	93.149	0.108
Full model	6.861	34.853	41.714	0.164

Implementation and process evaluation

The IPE results are organised thematically according to the five research questions. In each section, findings from the interviews and observations completed during the school fieldwork visits are presented alongside data from the post-intervention teacher survey conducted in June and July 2024 with a response rate of 70%.

IPE RQA How effectively does the Counting Collections training equip teachers to deliver the intervention and improve their understanding of key number concepts?

Overview of the Counting Collections training

Observations were carried out by the evaluation team at the first training day, which was delivered in person, and the subsequent sessions, which were online. The training includes a practitioner handbook and access to an online learning environment which provided opportunities for sharing practice and action planning. Training materials were clearly presented and accessible to participants afterwards, and the sessions included ample opportunities for professional discussions and reflection. The implementation and expectations of the four-part sequence of the Counting Collections session were clearly communicated and video examples of effective practice were included. Teachers reported that they felt well equipped to set up and deliver Counting Collections in the classroom. This was all reflected in the teacher survey and interviews, where feedback was very positive.

Teacher subject knowledge and the complexities of number development were an integral focus of the training. During training sessions, the online learning platform was utilised effectively to provoke further discussion and exemplification in a highly interactive way. Teachers were invited to post photographs of practice and share their insights into episodes of early number activity in their own classrooms. Very often these discussions were based on the focus of the training and highlighted significant activities, for example, recording or subitising. This was a particularly effective feature of the training as it combined teachers' experiences and practice with reflective discussion underpinned by the early number concepts being developed. The reflections and actions log was also a significant aspect of the training and allowed teachers to respond to any challenges as they arose. These challenges were noted by the developer, which often shaped the training content to meet the needs of the teachers and their ongoing questions and concerns about teaching number. The use of breakout rooms promoted insight through reflection and discussion.

Training to deliver the Counting Collections programme and its materials

The post-intervention survey, sent to all teachers delivering Counting Collections, showed high levels of satisfaction with the training and materials. Teachers particularly appreciated the whole day, face to face training: 69.2% rated this as excellent; 18.5% as very good. Online training sessions were also viewed favourably, with 49.2% or respondents rating these as excellent; 32.3% as very good. The virtual learning environment was rated by 70.4% of respondents as either excellent or very good. Respondents were also very positive about the online training materials, with 50.8% rating them as excellent and a further 40% as very good. Pupil resources received the highest rating of all elements of the training, with the Counting Collections boxes rated as excellent by 87.7% of respondents and the tools (for example, ten squares and other equipment to support counting and strategising) rated as excellent by 78.5%. Session logs, action plans, and shared galleries were also rated as excellent or very good by two-thirds of respondents and the learning trajectory documents were highly praised, with 69.2% of respondents describing them as excellent and 24.6% as very good.

Table 18: Perceived quality of training

	Excellent %	Very good %	Good %	Satisfactory %	Did not attend %	N
Whole day face-to-face training	69.2	18.5	4.6	0.0	7.7	65
Online training sessions	49.2	32.3	15.4	3.1	0.0	65
Virtual learning environment	39.1	31.3	23.4	4.7	3.1	64
Learning trajectory documents	69.2	24.6	4.6	1.5	0.0	65

Online training materials	50.8	40.0	9.2	0.0	0.0	65
Session log	32.3	36.9	26.2	4.6	0.0	65
Action plans	32.8	35.9	25.0	6.3	0.0	64
Shared galleries	40.0	33.8	24.6	1.5	0.0	65
Pupil resources (counting collections boxes)	87.7	9.2	3.1	0.0	0.0	65
Pupil resources (tools)	78.5	16.9	4.6	0.0	0.0	65

The positive teacher survey responses pertaining to the Counting Collections training were reflected in the semi-structured interviews that took place during school visits, where eight out of ten interviewees reported feeling well equipped to deliver the four-part routine effectively. Teachers commented on the quality of the training and appreciated its grounding in research and the practical, illustrative examples that were provided:

> *'The training was very good quality, and I think the way that it's delivered is easy to understand and easy to speak about to staff when we come back to school' (S3).*

> *'The quality has been good, and it's quite research led, and it is always backed up by examples and what it would look like in the classroom' (S7).*

Most teachers interviewed said that they valued the in-person training at the start of the programme as it allowed them to make connections with others in the group and gave them a confident start. Scheduling the face to face session at the beginning of the programme meant that teachers were able to build productive relationships before moving to online sessions. While seven out of 16 respondents stated that they would have preferred in-person training throughout the programme rather than online, they recognised that this might not be practical. Two teachers felt that online training was not always conducive to the work environment.

> *'I would have liked more in-person days, because that sometimes is easy to do an in-person training because you're interrupted in a morning. So, if you're in here you can quite often be called away or interrupted, whereas if you're out of the building…' (S3).*

> *'Definitely face to face. I found that much, much better to be honest. I mean obviously getting there is tricky but when you have to be out of class anyway, I feel like you may as well be face to face because it's more productive' (S9).*

The sharing of practice using video clips and the use of the online learning platform were seen as valuable training formats that reassured teachers and gave them an opportunity to share new ideas. Teachers also reported that they valued the discussion in breakout rooms, but some felt that this did not replace the in-person discussion experienced in the face-to-face training. However, teachers appreciated the support and responsiveness of the development team.

> *'I've really enjoyed it I think because but delivery of it has been good, and the support has been good, so we know that we can always contact them' (S3).*

Teachers also acknowledged the quality of the number-specific content in the training. Where teachers said that they already had some confidence, they found it useful to revisit the area and appreciated the time to look in-depth at early counting and number concepts. In addition, a focus on number was welcomed by respondents who felt that their training since becoming a teacher had often focused on other aspects of the curriculum, such as literacy.

Overall, the Counting Collections package of training and materials was well received by teachers responding to the survey, regardless of their prior experience. Those already feeling confident appreciated the time to study early number in more detail, and for those with less confidence and experience, their confidence and knowledge increased.

Overview of business as usual

Data from the fieldwork interviews highlighted that most schools held daily maths sessions spread across four or five days. Nine of the ten schools visited were using a mathematics programme or scheme. The most popular programmes

cited follow a mathematics mastery approach, including White Rose, Power Maths, Bemrose Maths, and Mastering Number. One school had recently begun the Karen Wilding approach while another used the Do It approach. The majority of schools said that they spent approximately 20 to 30 minutes a day on maths-specific teaching and learning. All schools stated that they adapted the programmes and schemes used to ensure that the more formal approach offered by the schemes sits within the less formal play-based approach used in early years. This comes through adaptations at both the planning and delivery stage. Over the last ten years, schools have been encouraged to use a mastery approach. There has been a nationwide push for publishers, consultants, and websites to take up the mastery brand (Boylan, 2025) resulting in a move towards the use of schemes and whole-school approaches, particularly in KS1 and KS2. This approach is now prevalent in most reception classes and has impacted on the structure of maths delivery. The National Centre of Excellence in the Teaching of Mathematics (NCETM) offers professional development to support schools to get the best out of whichever teaching for mastery scheme they use, whether that scheme is an 'off the shelf' commercial mastery scheme or one that the school has developed itself (NCETM 2024).

Data collected from the teacher survey shows that business-as-usual practice involving the use of maths schemes is well balanced across both intervention and control schools. Teachers from 100 of 130 schools reported using a scheme (48 intervention, 52 control). The most frequently mentioned was White Rose, with 32 intervention schools and 32 control schools reporting that they used it, followed by Mastering Number (15 intervention, 18 control).

The organisation and structure of the mathematics sessions followed a similar pattern across all of the schools. Standard practice in maths provision involves three main components: explicit whole-class teaching (often involving adaptations of maths schemes) followed with adult guided group work and then mathematics through the continuous provision. This refers to activities and resources that are available for children to access independently throughout the day, chosen to capture pupil curiosity and forming part of the enabling environment including outdoor learning areas.

> *'We will always have, like, a 10- to 15-minute, depending on the cohort, adult input. Whole class. Then we will pull small groups to do an activity with an adult. It might be the teacher some days and it might be one of the TAs, and then just lots of continuous provision throughout that time as well' (S8).*

> *'We have a carpet input adult led session and then we have a little bit of time on the carpet where we all do a bit of a practice, or a bit of an apply, and then the children get a learning challenge and then they do that within the provision' (S4).*

One school also had a mathematics intervention programme for individual pupils not meeting expected levels of attainment:

> *'So it would be about 15 minutes, 20 minutes for a whole-class teaching session a day, and then you would maybe do a 20-minute group activity a week with the children, and then obviously if they're choosing to … or there are certain children that you wanted to focus on, that you would encourage them to come to the maths area a little bit more, and then there are some children who are doing maths intervention as well as those that weren't at the expected level' (S6).*

All of the schools visited held a whole-class carpet session to mark the beginning of mathematics sessions, the content of which often came from the chosen scheme being used in school. Pupils then moved to continuous provision and group work, with the learning focus dependent on current learning needs. Teachers mentioned that they used ongoing formative assessment to gauge pupil learning needs. One teacher used RAG ratings to direct the learners to the correct provision and support:

> *'We normally RAG rate it, so I will have my class list, and then I will just highlight it: green—understood; orange—perhaps needs consolidation; and pink … and then I would also make notes. So it might be that the children that are your green children, they might need an extra challenge to push them on to the next—to sort of deepen that understanding. The yellow children might need a bit of a pre-teach or a post-teach' (S1).*

The teachers interviewed spoke of targeted, adult guided support either for individual pupils or small groups. This tended to link to the whole-class learning but with differentiated outcomes. Most teachers supported the groups needing further consolidation or greater depth and challenge. Where present, TAs tended to work with SEND pupils or those struggling to grasp the current learning outcome. Fieldwork observations and discussions with class teachers highlighted that this was not an expectation set by the intervention but a response to managing the differentiated outcomes for learners

during the four-part routine. Teachers felt that this allowed them to manage better the learning needs of all the pupils in the moment.

> *'You will see one of my TAs is working with those children just because they struggle to sit in a large class situation. They need to be in a smaller group. So, they will be separate and then I will have the other children all together with another TA supporting me' (S1).*

All of the schools mentioned that the explicit teaching was also supported by activities in the continuous provision and child-initiated learning brought through play.

> *'We always have maths opportunities. Yeah, we normally have at least two other activities … probably two other activities alongside the Counting Collections in our provision and they tend to be helping the children progress in their other maths session. So, they tend to be in line with these maths sessions, and we also do what is called a "linked activity"' (S5).*

All ten teachers interviewed felt that the programme fitted easily within the scheduled weekly maths as it followed the pattern of delivery already in place. Schools delivered the four-part routine at the beginning or the end of the week. Two schools that would normally have four maths sessions per week increased this to five when Counting Collections was introduced. Others substituted the Counting Collections session into the four- or five-day maths schedule:

> *'We used to do five and now we do four and one day of Counting Collections; so, it's five in total but one is Counting Collections' (S2).*

At most schools visited, Counting Collections sessions were ended by bringing the pupils back together to reflect on the learning. Teachers signposted specific examples of learning to consolidate and extend the pupils' mathematical understanding and to raise their awareness of potential learning for the next session. School visits showed that teachers also used this as an opportunity to give targeted praise (for mathematical thinking, recording, use of effective strategising).

It is worth noting that only three of the teachers interviewed spoke of recording being part of the weekly maths learning prior to using Counting Collections. Two of these mentioned using Power Maths books—a whole-school maths mastery approach that promotes the use of a maths book to record answers. This is specific to the programme: recording is not a usual practice in reception and is only mentioned briefly in early years statutory requirements.

Counting Collections and its impact on teacher confidence

The survey asked both intervention and control teachers to rate their confidence in teaching and supporting children with specific aspects of counting and how this had changed over the previous year. Intervention teachers were more likely to report improved confidence in this area. High numbers also reported major improvements in their confidence to teach the stable order principle (50% compared to 18.3% for control respondents), the cardinal principle (45.3% compared to 20% for control respondents), and the abstract counting principle (51.6% compared to 18.3% for control respondents). In addition, more intervention teachers reported major improvements in teaching and supporting pupils with verbal counting skills (35.9% compared with 5% in control schools). It is significant that teachers in control schools were more likely to say they were 'already fully confident' when asked about their confidence in supporting these aspects of counting and how far it had improved over the previous year. One possible explanation for this response is that without the Counting Collections training, teachers may not have been aware of the range of progressive steps required to scaffold learning and the skills associated with early counting. It is possible that involvement in the programme significantly raised teacher awareness of aspects of subject knowledge needed for teaching these specific concepts of early number and heightened their awareness of the steps of progression in children's mathematical learning.

A similar pattern emerged from the survey items on teaching subitising (Table 20). In intervention schools, major improvements were noted in the teaching of early numerosity (48%), perceptual subitising (57.8%), and conceptual subitising (56.3%). Control schools were less likely to report major improvements in these areas and more likely to report being 'already fully confident'. A possible explanation is that the targeted aspects of the training raised awareness of the complexities and significance of subitising. For example, the training focused specifically on the different types of subitising and promoted teacher awareness of how to use effective number interactions to support learners to apply perceptual subitising and the significance of this for making connections to early calculation. Most teachers interviewed

said that they were already aware of perceptual subitising yet conceptual subitising was a new concept to them, and the training had provoked reflections on how they could draw children's attention to this.

Table 19: Teacher confidence in supporting specific aspects of counting

Aspect of counting		No improvement: still lacking confidence	Some improvement	Major improvement	No improvement: already fully confident	N
Verbal counting	Con	0.0%	48.3%	5.0%	46.7%	60
	Int	0.0%	39.1%	35.9%	25.0%	64
One to one correspondence principle	Con	0.0%	32.2%	15.3%	52.5%	59
	Int	0.0%	39.1%	43.8%	17.2%	64
Stable order principle	Con	0.0%	41.7%	18.3%	40.0%	60
	Int	1.6%	37.5%	50.0%	10.9%	64
Cardinal principle	Con	1.7%	40.7%	20.3%	37.3%	59
	Int	0.0%	40.6%	45.3%	14.1%	64
Abstract counting principle	Con	3.3%	41.7%	18.3%	36.7%	60
	Int	0.0%	40.6%	51.6%	7.8%	64
Order irrelevance principle	Con	3.4%	47.5%	16.9%	32.2%	59
	Int	0.0%	40.6%	51.6%	7.8%	64
Counting in 2s verbally and counting objects	Con	12.1%	37.9%	6.9%	43.1%	58
	Int	1.6%	45.3%	40.6%	12.5%	64
Counting in 5s verbally and counting objects	Con	15.8%	35.1%	5.3%	43.9%	57
	Int	6.3%	43.8%	35.9%	14.1%	64
Counting in 10s verbally and counting objects	Con	14.0%	35.1%	5.3%	45.6%	57
	Int	3.1%	42.2%	39.1%	15.6%	64
Counting on (forwards or backwards from a given number)	Con	0.0%	42.4%	5.1%	52.5%	59
	Int	0.0%	43.8%	35.9%	20.3%	64
Counting using place value (counting in units including multiples of 100)	Con	19.3%	42.1%	5.3%	33.3%	57
	Int	7.8%	45.3%	35.9%	10.9%	64
Counting beyond 100 (verbally and with objects) using place value	Con	19.3%	43.9%	7.0%	29.8%	57
	Int	7.8%	46.9%	35.9%	9.4%	64

Table 20: Teacher confidence in teaching subitising

Aspect of counting		No improvement: still lacking confidence	Some improvement	Major improvement	No improvement: already fully confident	N
Early numerosity, to be able to distinguish between quantities	Con	0.0%	45.8%	18.6%	35.6%	59
	Int	0.0%	37.5%	48.4%	14.1%	64
Perceptual subitising	Con	0.0%	42.4%	25.4%	32.2%	59
	Int	0.0%	28.1%	57.8%	14.1%	64
Conceptual subitising	Con	1.7%	44.1%	25.4%	28.8%	59
	Int	0.0%	31.3%	56.3%	12.5%	64

The survey asked teachers to report on their confidence in teaching early number, including the planning and assessment of early number. These questions were structured to ask respondents whether they agree/disagree with a series of statements. The results are presented in Table 21. Most respondents (98.3% control, 93.6% intervention) either somewhat agreed or strongly agreed that they are confident in observing mathematics learning in continuous provision and identifying next steps in learning. Most respondents in both groups also somewhat agreed or strongly agreed that they are confident in supporting mathematics learning as it unfolds through high quality number interactions, however more intervention teachers (59.7%) than control teachers (49.1%) strongly agreed with this statement.

Teachers were also asked about their confidence in promoting mathematical thinking around number sense when pupils make spontaneous mathematics comments or discoveries. Again, most intervention teachers (96.6%) agreed that they were confident in this area, as did 93.6% of control teachers. Most respondents in both groups also agreed that they are confident in supporting children with number sense misconceptions, although more teachers in the intervention group strongly agreed (53.2%) than control (43.1%).

Teachers in intervention schools were more likely to strongly agree that they are confident in supporting children with mathematical paired discussion and learning (46.8% compared to 37.9% control). In both the intervention and control groups, most respondents (92%) agree that they are confident with this. Overall, 95% of teachers in both treatment groups agree that they are confident with supporting children by planning successful activities through the continuous provision to promote number sense, although there were marked differences in the number of teachers that strongly agree (59% in intervention schools compared to 34.5% in control).

More teachers in intervention schools (45.2% compared to 37.9% in control schools) strongly agree that they are confident in translating findings from assessment into curriculum plans. Again, overall confidence is high with 93% of teachers agreeing that they are confident with this. Mapping out the progressive steps of progress in number sense was another area where more intervention teachers strongly agree that they are confident (38.7% compared to 24.1% in control schools). Among control teachers, 22.4% did not agree that they were confident with this, compared to 11.3% for intervention teachers. Teachers in intervention schools are also more likely to strongly agree that they were confident in challenging and extending children's mathematical thinking through number sense (45.9% compared to 34.5% in control schools). More control teachers (12%) did not agree that they were confident in this area compared to counterparts from the intervention group (8.2%).

Confidence in supporting children with early number/counting and its link to early calculation was also very high, with 93.6% of teachers in intervention schools strongly or somewhat agreeing that they were confident in supporting children with this, whereas for control schools this was 93.1%. More teachers in intervention schools (53.2% compared to 44.8% in control schools) strongly agreed that they were confident in using effective modelling, scaffolding, and appropriate resources to support children’s understanding of the key concepts in number sense. Most teachers (93%) agreed that they were at least somewhat confident. In intervention schools more teachers strongly agreed that they are confident in drawing out children’s mathematical understanding from their jottings and recordings (45.2% versus 27.6% in control schools). More control teachers (12%) did not agree that they were confident in this area compared to counterparts from the intervention group (6.4%).

Table 21: Teacher confidence in teaching early number

		Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	N
Observe maths learning in continuous provision and identify next steps	Con	0.0%	0.0%	1.7%	41.4%	56.9%	58
	Int	3.2%	0.0%	3.2%	35.5%	58.1%	62
Support maths learning as it unfolds through high quality number interactions	Con	0.0%	1.8%	3.5%	45.6%	49.1%	57
	Int	3.2%	0.0%	3.2%	33.9%	59.7%	62
Promote mathematical thinking around number sense when they make spontaneous math comments/discoveries	Con	0.0%	1.7%	1.7%	48.3%	48.3%	58
	Int	3.2%	0.0%	3.2%	41.9%	51.6%	62
Support children with any (number sense) misconceptions	Con	0.0%	0.0%	1.7%	55.2%	43.1%	58
	Int	3.2%	1.6%	3.2%	38.7%	53.2%	62
Support children with their mathematical paired discussion and learning	Con	0.0%	1.7%	5.2%	55.2%	37.9%	58
	Int	3.2%	0.0%	4.8%	45.2%	46.8%	62
Plan successful activities through the continuous provision to promote number sense	Con	0.0%	1.7%	3.4%	60.3%	34.5%	58
	Int	3.3%	0.0%	3.3%	34.4%	59.0%	61
Translate findings from assessment into curriculum plans	Con	0.0%	3.4%	1.7%	56.9%	37.9%	58
	Int	3.2%	0.0%	4.8%	46.8%	45.2%	62
Map out the progressive steps of progress in number sense	Con	0.0%	8.6%	13.8%	53.4%	24.1%	58
	Int	3.2%	1.6%	6.5%	50.0%	38.7%	62
Challenge and extend children’s mathematical thinking in number sense	Con	0.0%	3.4%	8.6%	53.4%	34.5%	58
	Int	3.3%	1.6%	3.3%	45.9%	45.9%	61
Support children with early number/counting and its link to early calculation	Con	0.0%	1.7%	5.2%	43.1%	50.0%	58
	Int	3.2%	0.0%	3.2%	46.8%	46.8%	62
Use effective modelling, scaffolding and appropriate resources to support children’s understanding of the key concepts in number sense	Con	0.0%	0.0%	6.9%	48.3%	44.8%	58
	Int	3.2%	0.0%	4.8%	38.7%	53.2%	62
Draw out children’s mathematical understanding from their jottings and recordings	Con	0.0%	6.9%	5.2%	60.3%	27.6%	58
	Int	3.2%	0.0%	3.2%	48.4%	45.2%	62

These trends were also apparent in teaching aspects of counting (Table 22). However, the most common response here was to report ‘some improvement’ in confidence. Intervention teachers are more likely to report major improvements than their control group counterparts, and control teachers are more likely to report being already fully confident. One possible explanation for this is that the Counting Collections training raises teacher awareness of the components and early number skills required to teach early counting effectively.

Table 22: Teacher confidence in teaching aspects of counting

Aspect of counting		No improvement: still lacking confidence	Some improvement	Major improvement	No improvement: already fully confident	N
Different ways to partition a whole number. Identifying pairs of numbers that make a total, exploring the idea that whole numbers can be split into two groups or more	Con	1.7%	39.7%	19.0%	39.7%	58
	Int	0.0%	44.4%	42.9%	12.7%	63
Decomposing 10, to begin to develop number facts and known number bonds that will build the foundations of understanding number	Con	3.4%	37.9%	22.4%	36.2%	58
	Int	0.0%	47.6%	42.9%	9.5%	63
More than and less than	Con	1.8%	35.1%	19.3%	43.9%	57
	Int	3.2%	49.2%	31.7%	15.9%	63
Patterns within numbers up to 10, including evens and odds, double facts and how quantities can be distributed equally	Con	3.4%	41.4%	17.2%	37.9%	58
	Int	1.6%	44.4%	42.9%	11.1%	63

Table 23 shows teacher survey responses in teaching aspects of numerosity/comparison. The patterns illustrated above are again apparent, with control teachers more likely to say that they were already fully confident, and intervention teachers more likely to report major improvements in their confidence. For most items, the number saying ‘some improvement’ is similar between intervention and control schools, however, control teachers were more likely to respond with ‘no improvement—still lacking confidence’ for aspects such as comparing using one to one correspondence (3.4%), perceiving differences (8.6%), and the numerosity of the comparison (3.4%). This may be due to the opportunities provided by the training to reflect on the mathematical concepts in practice, which allowed teachers to challenge their thinking around the relevance of numerosity and comparison in relation to a progressive model of teaching and learning. Data from the teacher interviews highlighted that the introduction of the learning trajectories were instrumental in helping teachers to look for these mathematical connections and progressive pathways: teachers may well be confident in their subject knowledge while still not fully appreciating the depth and complexity of that subject knowledge in practice—something that is only revealed when it is highlighted. Shulman (1986) calls this ‘pedagogical content knowledge’.

Table 23: Teacher confidence in teaching aspects of numerosity/ comparison

Aspect of counting		No improvement: still lacking confidence	Some improvement	Major improvement	No improvement: already fully confident	N
Sensing sameness (May use one to one or many-to-one correspondence in certain situations)	Con	1.7%	39.7%	19.0%	39.7%	58
	Int	0.0%	44.4%	42.9%	12.7%	63
Sensitive to the relation of ‘more than’ and ‘less than’ involving very small numbers.	Con	3.4%	37.9%	22.4%	36.2%	58
	Int	0.0%	47.6%	42.9%	9.5%	63
Matching and sharing	Con	1.8%	35.1%	19.3%	43.9%	58
	Int	3.2%	49.2%	31.7%	15.9%	63
	Con	3.4%	41.4%	17.2%	37.9%	58

Comparing using one-to-one correspondence	Int	1.6%	44.4%	42.9%	11.1%	62
Perceiving differences	Con	8.6%	48.3%	15.5%	27.6%	58
	Int	0.0%	55.6%	34.9%	9.5%	62
The numerosity of the comparison, i.e. the number of things not the size of them	Con	3.4%	55.2%	8.6%	32.8%	58
	Int	1.6%	60.3%	30.2%	7.9%	62
Small ordinal numbers	Con	1.7%	39.7%	13.8%	44.8%	58
	Int	0.0%	49.2%	30.2%	20.6%	62
Comparing similar items	Con	1.7%	41.4%	15.5%	41.4%	58
	Int	0.0%	46.8%	37.1%	16.1%	62
Estimating numbers to 5 on a number line	Con	1.7%	50.0%	15.5%	32.8%	58
	Int	0.0%	51.6%	38.7%	9.7%	62
Estimating using space and number	Con	1.7%	46.6%	15.5%	36.2%	58
	Int	0.0%	48.4%	40.3%	11.3%	61

To summarise the results presented in this section, comparing the self-reported confidence of intervention teachers and those in the control group suggests that Counting Collections has a positive impact on teachers' confidence in the following aspects of number:

- mapping out the progressive steps in number sense;
- supporting pupils with misconceptions;
- holding high quality number interactions with pupils;
- extending pupil mathematical thinking through number sense;
- modelling and scaffolding, using appropriate resources;
- drawing out mathematical understandings from children's jottings and recordings;
- sensing sameness;
- more than/less than;
- comparing amounts;
- using one to one correspondence; and
- estimating using space and number.

Delivering Counting Collections

Most survey respondents (97%) reported delivering Counting Collections as a standalone session. Regular features of the session involved arranging pupils into pairs to share their collection of objects and carry out the specified four-part routine (94%). Most teachers modelled a specific approach or extended children's thinking as part of the four-part routine (79%). In many cases, the teacher introduces a counting focus to the class to begin with (68.7%) then draws the class together at the end of the session, choosing examples of pupil counting strategies to share with the class (67.2%). Many sessions were typified by the teacher observing and interacting with the children as they engaged in paired discussion, counting, strategising, and representing the structure of their count on paper, with 65.7% of survey respondents agreeing

that this was part of how they delivered the intervention. It is clear from the survey that the four-part routine was maintained in most sessions.

Observational findings showed that the teachers maintained the four-part routine but shaped the delivery to line up with the organisation of the classroom and practices. The choosing part of the routine showed varying approaches and practices. Some teachers let the pupils choose, others guided their choice, and some gave them a specific box to promote an appropriate level of challenge. Children moved through the different parts of the routine with ease and knew the expectations for each part.

Fieldwork observations of the four-part sessions highlighted that the number interactions between teachers and pupils were of a high quality and that the teachers demonstrated secure subject knowledge, enabling teachers to identify key learning moments and use effective number interactions to consolidate number learning**.** This was most obvious when children were transitioning through the five principles of counting and securing their counting development. These findings highlight the perceived positive impact of the Counting Collections training and delivery on teacher subject knowledge.

Fieldwork observations showed that teachers knew how to plan and teach weekly, whole-class, four-part Counting Collection sessions. Most teachers followed similar delivery practice, starting each session with explicit, direct teaching on the carpet. The teaching observed was high quality and set the tone and the learning expectations for the four-part routine. Carpet sessions often involved visual resources, models and images, and active engagement. Fieldwork observations noted that carpet sessions usually involved a specific focus that the teacher wanted to highlight and an opportunity to consolidate understanding. Most teachers chose the focus of the session based on the previous session, and any area of learning that they felt the pupils would benefit from practicing further, therefore drawing on a progressive model of teaching and learning. Alternatively, teachers used the explicit teaching part of the session to introduce a new concept that they wished to bring out through the Counting Collection routine, such as estimation. Other teachers modelled the four-part routine focusing on the part of the session that the pupils needed extra support with, for example, strategising or recording. A focus on strategising was evident across most settings, perhaps due to the timing of the visits and the fact that this was an emerging need in establishing the programme. Observations showed that the focus on strategising was skilfully managed and based on a secure understanding of the potential challenges of this part of the routine, but also on its potential for developing mathematics learning.

Observations highlighted many examples of good practice and skilled number interactions including the use of a mini-plenary to address a common problem with children using the five frame for larger numbers to point out the inefficiency and to highlight the larger ten frame. The children were encouraged to check and review their work. A second teacher was observed demonstrating a misconception, how another pair had used mirrors to count but they covered four squares on the ten frame with one mirror. The teacher pointed out the problem and suggested finding something smaller to count. A third teacher noticed that two pupils had arranged their large counters in fours and she said three rows of four is 12 and then followed up with, ‘So how many?’, leading the child to respond to the correct answer.

Observations found that most teachers finished the Counting Collections session by bringing the class back together on to the carpet and reflecting on the learning that had taken place. The teachers used this to consolidate and challenge the pupils’ thinking ahead of the next session. This part of the session was considered a strength by the observers, promoting a positive attitude to learning through the four-part routine and celebrating the pupils’ efforts.

During fieldwork observations, teachers gave frequent reminders about the required organisation and classroom management of Counting Collections sessions. The atmosphere in the classroom across all schools was positive and conducive to learning.

IPE RQB Is the intervention associated with improvements in teacher knowledge of learning trajectories?

The learning trajectories (LTs) component proved to be one of the most valued aspects of the programme. As discussed above, the LT documents provided as part of the programme were well received by teachers, with 69.2% of survey respondents describing them as excellent and a further 24.6% as very good. Teacher interviews conducted during fieldwork visits also showed favourable views toward this element of the intervention. Previously, teachers had relied upon the Early Years Statutory Framework to assess children and look for next steps in learning. Teachers reported that

they had used LTs for training other staff and for other interventions to support assessment, and one had used them during parent consultations. One teacher was interested in exploring further how the LTs might be used to support the pupil's journey towards the Early Learning Goal in maths. Teachers stated that the trajectories were being implemented in different ways in school, and appreciated the incremental steps that children take in understanding number.

> *'I have never really looked at the trajectories similar to the ones that have been provided, so again that has been a real eye-opener of the different steps that the children should take, and the way that it progresses through, and I think that has been really useful' (S1).*

LTs were useful in supporting additional adults to support Counting Collections, and for staff to share assessment information about individual pupils:

> *'The TAs have access to the learning trajectories but, also, we quite often have a biweekly meeting just to talk it through and in there we look at the trajectories together to see what we need to do. Learning trajectories have been really useful because we have them in the classroom and the additional adults access them as well and it's easy for us to pick up the learning trajectory and know how to move a child on' (S3).*

It is clear that teachers believed their knowledge of learning trajectories to have been improved by participating in Counting Collections. Teachers commented that they would continue to use them in their maths teaching after the trial. Teachers also expressed an interest in how they might be tied in with statutory expectations as an additional assessment tool, while others felt this would be replicating what they already have in school. Future iterations for the programme may consider how this resource might be used further.

IPE RQC Do teachers believe that the intervention is associated with short term improvements in pupil confidence, enjoyment, and attainment in number?

In the post-intervention survey, teachers from intervention schools were asked how they perceived the impact of Counting Collections on pupil enjoyment, confidence, and enjoyment and various aspects of maths learning deemed central to the programme (Table 24). The greatest impact was 'working collaboratively' (77%). Recording mathematical thinking and ideas was another area where teachers believed the programme had significant impact (76%). Teachers also noted significant impact on pupil mathematical talk, enjoyment of mathematics, and number confidence (71%). The area where teachers were least likely to report significant impact was pupil ability to concentrate for longer on maths learning or tasks but almost half of the sample (49%) noted a significant impact on this area.

Table 24: Impact on pupils

	Significant impact	Some impact	No impact	Hard to say	N
Working collaboratively	77.3%	21.2%	0.0%	1.5%	66
Recording mathematical thinking/ideas (representing the structure of their count on paper)	75.8%	22.7%	0.0%	1.5%	66
Mathematical talk	74.2%	22.7%	0.0%	3.0%	66
Enjoyment of mathematics	71.2%	24.2%	3.0%	1.5%	66
Number confidence	71.2%	27.3%	0.0%	1.5%	66
Learner confidence/positive attitude towards maths	68.2%	30.3%	0.0%	1.5%	66
Problem solving skills	60.6%	36.4%	0.0%	3.0%	66
Exploratory talk (open ended communication)	54.5%	39.4%	1.5%	4.5%	66
Working strategically	54.5%	43.9%	0.0%	1.5%	66

Metacognitive talk (thinking out loud)	51.5%	43.9%	0.0%	4.5%	66
Pupil ability to concentrate for longer on maths learning/task	49.2%	49.2%	0.0%	1.5%	65

Teachers were also asked which of the four parts (choose, strategise, count, record) had the most impact on pupil number sense. They were least likely to say that 'choose' impacts positively with only 31% of teachers agreeing with this. At least three-quarters of teachers (76%) said that 'strategize' had a positive impact. Most teachers said that 'count' (82%) and 'record' (85%) had a positive impact. Interview data highlights that teachers felt that the 'choose' part of the sequence was sometimes problematic. The collections were differentiated by the number of objects and were colour-coded. Teachers expressed some concern about children choosing a collection that did not align with their current learning need. An added challenge was that children choose their collection in pairs. This may explain why only 31% felt that the 'choose' part impacted positively.

Often teachers described not routinely giving children opportunities to record prior to Counting Collections, which may explain the perceived the impact of this part of the session, as children's mathematical thinking was made 'visible'. The 'strategise' part of the sequence was reported to be the most challenging, but also an area of improvement tied to mathematical thinking, mathematical talk, and problem solving. The teachers felt that they had gained most from this area in terms of noticing pupils' current thinking. It may be that the teachers focused more on this than the impact of this part of the sequence on strengthening skills in number sense. Fieldwork findings reiterated that the areas of greatest impact identified by the teachers were the children's ability to count and to record.

In fact, across the fieldwork observations and the teacher survey (Table 21), the data showed that children's representation of number was a significant feature of the programme. Teachers interviewed reported that the greatest impact had been on mathematical thinking and mathematical recording. Data from the teacher survey showed that 45.9% felt more confident in challenging and extending children's mathematical thinking through number sense compared to 34.5% in control schools. More intervention teachers strongly agreed that they are confident in drawing out children's mathematical understanding from their jottings and recordings (45.2% compared to 27.6% in control schools).

Interviews with teachers suggested that the intervention impacted on six main areas: intrinsic motivation, concentration, working cooperatively, mathematical talk, recording mathematical thinking (jottings, giving meaning to mark making), and number confidence. Most teachers felt that the greatest impact was on mathematical recording**,** which was seen by some schools as unimportant prior to their involvement in Counting Collections. The routine of Counting Collections was also seen as supporting independence, communication, and collaboration between pupils.

Interestingly, teachers very often linked their own growing confidence teaching the Counting Collections routine with children's growing interest in mathematics confidence:

> *'Their attitude, they absolutely love … they are a brilliant, brilliant class. They love it. They love maths, they love when it's CC [Counting Collections] and they love telling other people that they're doing CC that day and like when we do our calendar. It's a little bit different' (S10).*
>
> *'I think that it's definitely changed [my practice] because at the start there is a lot of resources and it was trying to learn how best to put it out for the children and also in the lesson, how to organise it. I've tried it on tables, I've tried it on the carpet. I went from the carpet to the tables, but then the children were so spread out I found it harder to get round, so I brought them back to the carpet. I think that now they've gotten more confident in what they have to do it's easier. Because at the start it's modelling each step, whereas now I can just ask them what do we need to do, and I maybe put pictures up as prompts and that is something I've started doing' (S4).*

High levels of pupil engagement during sessions appeared to be in part attributed to the collections, and in part attributed to working with a partner. Pupils observed by the evaluation team during the school visits were enthusiastic about the session. Observations and teacher interviews showed that two key factors were most evident in the success of the programme with regard to pupil impact and attitude to learning: the children were intrinsically motivated and showed high levels of concentration over the session. Pupils enjoyed the concept of counting a collection of items and although

some teachers would have preferred to guide the choice of the collection, the opportunity for pupils to choose their own collection appeared to drive their enthusiasm.

Most pupils worked in pairs to discuss the strategy for counting the collection. Observations showed that this was the area they found most challenging, yet it promoted mathematical thinking. It was observed that this part of the routine promoted a depth of thought about the quantity of items they were counting and the best approach for that amount. In children having to evaluate their choices and the efficiency, it provided key learning points around counting on, the quantity, early numerosity, and ordinality. The most effective practice involved teachers supporting pupils to reconsider the resource used. For example, one pair had their collection arranged in a round dish, which caused some confusion. This strategy was discussed with the teacher, who suggested a different resource such as the tens frame.

Teachers noted that pupils enjoyed many aspects of the sessions, including choosing the collection, trying different ways of counting, and recording. This was seen as leading to improved engagement with maths and excitement about the subject. Some also explained how higher levels of engagement and motivation impacted positively on pupil counting skills:

> *'We have found that they're more engaged with maths, so that they're more excited by maths, and that their counting skills have improved. They were very good at counting anyway to 20 but counting beyond the 20 they always struggled. And counting backwards and forwards' (S3).*

> *'They are motivated by the objects and, as I say, because we don't have them all out all the time, as soon as I refresh it. So, I refresh it every two weeks and I've got a cupboard in there with the rest of them in and we change it around and they do like when it's all changed around' (S6).*

Teachers interviewed felt that the impact on pupil social skills and attitudes towards learning aided their participation and gains in number confidence. One said that the class was now doing more paired work having taken part in Counting Collections. This was sometimes described as difficult to begin with, as pupils would try to start counting on their own, but the collaborative element was seen as ultimately benefitting pupils, with one teacher saying that working with a partner had the greatest impact of the programme in terms of mathematical thinking and problem solving. Other respondents highlighted that the pair work was significant in raising the quality of mathematical talk. Furthermore, teachers were keen to point out how pupils had grown in independence, confidence, and skill in collaborating and communicating through following the Counting Collections routine.

Children's recordings

There was considerable evidence from both fieldwork observations and teacher interviews to suggest that Counting Collections has a positive impact on mathematical recording. Two teachers commented that this was something they would not have thought about prior to the programme. Practitioners also commented on the noticeable progression from pictorial to iconic, and then on to symbolic representations and notations alongside more efficient mark making and writing numerals:

> *'I think the recording element is the massive part, because not only are they counting, and then they've got their amount, but then when they record, they are counting again because either they're doing dots or lines or whatever. And then now they are even starting to write the numbers as well, which wouldn't have happened in another year' (S5).*

> *'Recording is something that we maybe hadn't focused on as much before when we were recording numerals, and they surprise you. I was very surprised that the little girl who wrote numbers to seven, I had never seen her do that before' (S9).*

Most teachers reported the recording component as being a valuable part of the programme. A contributing factor to the learning success linked to mathematical recording was perhaps that it was the key component of the four-part routine that most challenged business as usual practice. Mathematical recording is not prominent in statutory expectations until KS1 and is sometimes overlooked. The intervention delivered high quality training that had a focus on mathematical recording and mathematical jottings, raising teacher awareness.

Impact on early number skills

During interviews, practitioners described how the following aspects of early number had progressed through the Counting Collections programme: counting objects more confidently, the one-to-one principle, numerosity and cardinality, number comparison, and conceptual subitising.

> *'They are starting to see patterns in numbers, awareness of composition of number, their counting ability is, you know, amazing and they are more aware of the position of the numbers within the 100 square I think, and within the number line' (S8).*
>
> *'He had lots of different Numicon bits on his place, and he was saying 3+3 = 6 and 1+1+1+1=4. And he had done that because he's explored that within Counting Collections and he has made that awareness of composition of number' (S8).*

Two teachers commented on the ability of pupils to see patterns more readily. Three teachers described how problem solving during the strategise part of the session had had a noticeable impact on number confidence and mathematical thinking.

> *'I think the learning to strategise and problem solving is having the most impact. It has helped them to think mathematically, and problem solve' (S3).*

However, teachers also acknowledged that for some, strategising was the most challenging aspect of the four-part routine and this needed a lot of modelling. Observations and discussions with teachers in schools highlighted that this was often dependent on pupils' understanding of the tools they had chosen and how to use them. This was sometimes hindered by the pupils' ability to communicate their mathematical thinking. This was most noticeable with pupils who were struggling with the basic rudiments of counting. However, there was evidence of good practice where teachers facilitated quality interactions to help promote pupil mathematical thinking by modelling how to use the tools they had selected to count the collection.

> *'"How should we count it? I said how shall we count it?" One of them just went, "1, 2, 3, 4 . . .", and I said, "I can see you're counting them already, but how are we going to do it? Let's talk to our partner." So it is trying to encourage the different ways of counting using the resources, that is why I tried to go around all of them. I've started going around all of them and supporting that conversation and then, when they've decided, I go, "Okay, now go and get the resource"' (S4).*

The consensus was that the programme had impacted learning positively and was worthwhile:

> *'Just how independent they are and how accessible it is for all children. And I just found it fascinating. You really get to understand how your children are as thinkers, and learners. So we have just found it really, eye-opening' (S1).*
>
> *'It just works. It is fun. The children love it as much as the staff. They get so much out of it and sitting back and watching their development over the last few months has been wonderful' (S8).*

It is clear from the survey and interviews that teachers in intervention schools perceived Counting Collections to have had a positive impact on pupil engagement, enjoyment, and maths. It was also seen as improving pupil concentration and ability to focus on tasks for longer. A significant impact was centred around pupil jottings and informal recordings to represent counting and mathematical thinking. Pupil collaboration skills improved and talk about early number became more sophisticated. This aspect of the programme was welcomed by teachers, and it also provided a window (along with the opportunities for paired mathematical talk) to gain a deeper awareness of pupil understandings in early number.

Higher attainers

Another key theme that emerged during classroom observations was linked to challenge for the higher attainers, who were moving beyond counting. To meet the learning needs of higher achievers, one school was directing the learning to the composition of number and early place value through the modelling of tens frames and a number square. There was support with 'estimate' alongside guess, and one teacher modelled counting in tens as well as ones. They used place value to draw attention to the position of a number on the 100 square and how close numbers are to each other by counting on. There were a few examples of comparison and composition of number with higher attainers, but this

was a minority. Overall, teachers stated that they would like more information and training on extending higher attainers through the Counting Collections routine.

Lower attainers and whole-class mathematics teaching

Some tensions expressed in teacher interviews reflected a broader concern over whole-class maths teaching and meeting the learning needs of all pupils. Teachers developed strategies to try and ensure that they offered all pupils an opportunity to usefully participate. They also expressed concerns that managing the session could be a challenge when some children finish the task and others are taking longer to complete their recording.

> *'Like I had a TA today but that is not always the case and when I'm doing it on my own I feel like your really low ability children are just kind of sitting there. And you can't get around them all because of time constraints and whatnot' (S9).*

> *'Normally it might just be me. There is not often a teaching assistant with me at that point but if they are then they're usually supporting the lower ability. And I try and get around them all. You can't always get and check everybody, and you definitely couldn't if you had the whole class. I can with most of them but I can home in on some' (S6).*

The above quotations from interviews with teachers reflect the challenges faced and highlight the crucial role of the TA in the delivery of whole-class maths lessons. It also reflects the wider concern of teachers in needing to address misconceptions at the point of learning and meeting the learning needs of all pupils.

Another contributing factor appears to be around the structure of the four-part routine and, specifically, teacher concerns about how to provide targeted support for the lower attainers through the different stages of the four-part routine in the context of a whole-class approach. In the absence of a TA, the learners often struggled to move through the four parts of the routine without assistance, particularly with strategising and recording. Some teachers tried using mixed ability pairs to help with this but reported that this often caused issues. This also appears to be linked to the choice of collections: teachers commented that the different collections offered up different learning opportunities and this was difficult to oversee in a whole-class approach where pupils chose their own collection.

> *'So, then I grouped them by ability and that was better and I liked it much better when I gave them specific coloured boxes to choose from. So, I knew that they were counting at the level that they needed to be counting at' (S6).*

The above quotations highlight the challenges of the whole-class approach and the management of the different elements that make up the four-part routine to ensure appropriate differentiation at the point of learning.

In the context of the Counting Collections programme, differentiation comes through the teacher's number interactions with the learners as they are strategising and counting the learning needs as they unfold. Based on the views of the teachers this poses significant challenges for them when the whole class is accessing the programme at the same time. Teachers felt a significant pressure to support all the emergent mathematical development of all the learners through deliberate practice to help avoid any misconceptions. This was not an expectation set by the programme, but teachers were concerned about pupils not making progress because they could not get around all the children during the session. In the absence of a TA they worried that the lower attainers would be most affected by this.

IPE RQD What level of fidelity is observed during the trial and what influences fidelity?

Access to the Counting Collections library

Use of and access to the Counting Collections library is a key feature of the programme. Visits to schools confirmed that all classrooms had a collections library that was well resourced and situated in a communal area near a carpet or table. This was accessed at different times during the day in different schools. Most chose to oversee the pupils' use of this area rather than pupils accessing this through free play. The Counting Collections library was made available for pupils to play with during time allocated to continuous provision. Observations of children accessing the library outside of the weekly routine showed that most did so alone and there was a tendency to transfer the four-part routine and skills into the provision. Teachers reported that there was not as much uptake in the Counting Collections library as they would

have liked, and it tended to be the same pupils that accessed the provision. This was often the higher attainers who felt confident in working more independently on their own using the four-part sequence. As a result, some teachers gave access to the provision on certain days to encourage its use, for example, after the Counting Collections session. However, some low-attainers may have not wanted to continue working with the collections after the session due to cognitive overload, preferring alternative play.

Other teachers said that they had encouraged pupils to play in the library on the days that they did not do the four-part routine but there was little uptake. This may be due to the pupils seeing the resources as being part of the structured learning time rather than a free play activity.

In interviews, some teachers said that pupils initially accessed the Counting Collections library but, as items were sometimes misplaced, they had limited access to using the collections on a nearby table or carpet area at set times. Teachers had developed rules to make this manageable. For example, in one school, pupils have to ask to play with the Counting Collections library, they can only have three boxes out at a time, they are not allowed to bring the objects elsewhere in the room, and they have to return the resources to the library once they have finished with them. The restricted access may have had a detrimental effect on the potential of the library as a continuous provision to promote incidental learning for all children,

On the occasions where pupils were observed accessing the library the adult interactions were of a high quality and supported progress in early number sense.

The four-part routine

The four-part routine of the session was maintained during the trial. In most cases, the Counting Collections session was taught as a discrete session. There were a number of factors that could potentially have influenced fidelity to the programme but, as detailed in the section below, these were often resolved by teachers as they worked through and understood the programme more thoroughly. Two challenges emerged: the balance between child-led and adult-guided learning and, related to this, the role of the teacher in differentiating and extending mathematical learning.

While teachers described positive changes in their subject knowledge, they also articulated a series of tensions relating to programme delivery. These centred on striking a balance between child–led and adult guided learning, and the teacher's need to differentiate learning for individual children. Teachers reported initially feeling at odds with the structure in terms of choosing part of the routine. The collections are colour coded and vary in size. Pupils are encouraged to choose their own box. Some teachers raised concerns that pupils often selected one that was too easy or too difficult. Teachers reported that one of the main challenges was around grappling with the idea that differentiation comes through learning outcomes that are determined by the collection chosen by the pupils. These tensions were expressed as follows:

> *'Part of the programme is them having the choice and without taking that choice away, you know, it's great because they're all colour coded so you can steer them in that. I suppose the differentiation is tricky to get my head around' (S2).*

> *'I would have quite liked to have started that a bit earlier and not had all the larger amounts out, and maybe only introduced them later on. Are they just constantly learning a misconception? I don't know. I do struggle a bit with children who choose boxes that are too many and however much I try and say, obviously they are allowed to choose whatever but like. So how to tailor . . . I've taken a lot of the boxes away because to start with we had far more boxes and so now I just sort of have the majority red, just because that's where we are and I've got a few of the yellow and blue for an extension, and then I rotate those around every three weeks to put different manipulatives in' (S2).*

Some interview responses revealed a concern about teachers not being able to support all children during the session: moving around each pair in turn and not having the opportunity to address misconceptions was problematic:

> *'One of the things that I have found that I really struggle with that's been an issue is that because it's such a large group and there is often very few members of staff, sometimes it's just me doing it, and if I'm not able to be there to kind of affect misconceptions, those misconceptions are keeping happening' (S10).*

Despite the tensions raised, fieldwork observations and interviews showed that most teachers used various adaptations to successfully alleviate this and became more aware of the scope for adaptation of sessions as they became more confident with the programme. Strategies included thinking about pairing pupils with those of similar or different ability, the organisation of the collections, and directing the choice of the collection:

> *'Mixed ability pairs, and I did try that to start with but I just found the lower ability were not able to have a go and the higher ability weren't counting big enough amounts because you were having to tailor it so that the lower ability could still manage a little bit of this and the higher ability took over the lower ability. So, then I grouped them by ability and that was better. I then chose their boxes for them' (S6).*

> *'They really like the different collections, and we have had arguments over who is going to have [a particular one], and we try to change them so that they don't get bored. And also, they don't always pick the same one. I'm trying to say, "Can you pick a different one? I think that you counted that last time, and the time before."' (S1).*

Adaptations from one teacher included more explicit teaching at the start to guide pupil attention towards specific learning intentions. Another reported that she had gradually brought in more explicit teaching at the beginning and end of the session and had chosen the partners that they worked with, putting them into ability pairings.

Teachers reported that some pupils were beginning to explore the partitioning of larger collections and were keen to develop the potential learning opportunities around the relationship between partitioning leading to early place value. Teachers reported that the training around composition and decomposition of number was most helpful but would have liked further content on how counting skills might support early calculation:

> *'So, you know what I have wanted to do before is look at how we could extend it into those other sorts of skills like sharing, adding, subtracting, one more, one less' (S9.)*

Five of the ten teachers interviewed stated that they would have liked the training to cover how to extend and challenge more able learners. Fieldwork observations found that some teachers were less confident in number interactions with more able learners. They reported challenges in delivering appropriate levels of challenge within the four-part routine, particularly for pupils who were already proficient in composing and decomposing numbers. This limited the opportunity to extend their learning by making connections to structural understanding that supports early addition and subtraction. A few teachers stated that they were unsure how to bring challenge through the choice of tools and the strategies for counting larger collections. Responses from the teacher interviews showed that, for some, the training and the delivery of the programme had made them think more carefully about the higher achievers.

> *'When I was modelling the number 36, I was trying to get them to see it in tens, so then you can see the progression of counting, especially for the highers—counting ten, there is ten there, twenty, thirty. I don't need to count them again and know there's thirty there. So, you can see that could help. And one of the things that the trainers have said in the last session was actually they might get to the point where instead of drawing 30 counters, if you've got 36, instead of drawing 30 counters they might just put 10 and 10, write the number ten, and write the number ten again. And then they can see it's all the place value, isn't it, and partition it into tens, or three tens and a six. And that is going to really help them, when they go into Y1' (S6).*

The challenges teachers reported in the survey and interviews did not become an issue of compliance, which was achieved in the majority of schools.

What do the trial findings indicate about scalability?

Teachers were asked how they found integrating Counting Collections into their practice, and whether they might continue with the programme after the trial. The training and resources were highly valued, and this seemed to be an important factor for teachers and leaders deciding whether to continue. Most teachers wanted to continue but they also wanted to keep the resources and requested more content around the collections and how these might be used for challenging pupils and taking them forward to early calculations. The support, guidance, and training of the development team was also a significant factor when deciding whether to continue with the project. One school leader said:

> *'I think that the issue that you do have is when you're implementing something new: there is a lot of staff training involved and there is a lot of resourcing, and I think that when we took part in this trial, obviously, we were given all of those resources, and we had it all there. If you were a school that was looking to implement that yourself, there is a lot of work that would go into it and a lot of staff time, and I think that is how that would be managed. I don't think that schools have all of that kind of thing readily available in such a nice way, so I think that it's the resources side of things I would think is something to consider for other schools' (S1, maths lead).*

Schools were asked during field visits if they are likely to continue with the programme; one (of the ten visited) was keen to roll the project out in nursery; another wanted to continue into Year 1; some were thinking about how they could continue with Counting Collections amid other initiatives and maths programmes in school.

The fieldwork data shows that the programme is associated with perceived improvements in teacher subject knowledge in mathematics, and in pupil learning and attitudes in mathematics. Other impacts include pupil collaboration and independence. In particular, the value of recording and mathematical talk, and its influence on mathematical thinking and engagement in maths learning, was evident. The learning trajectories that are a key feature of the programme were seen to be highly valued by teachers and were being integrated into school provision.

Counting Collections is a programme that can be integrated easily into practice and is generally seen to complement business as usual in intervention schools. This is in part due to the fact that whole-class mathematics teaching is already established practice in many schools. As a result of this, and the reported positive impacts and schools' interest in early number training, it has significant potential for scalability. Certain aspects of the programme, such as the focus on recording, communication and collaboration, and learning trajectories, were significant features that teachers and leaders interviewed welcomed. Suggestions for further development of the programme involve a strategy for supporting lower-attainers and also guidance on the ways in which children progress from counting to strategising across the session.

Cost

According to information provided by the programme developer, a school seeking to deliver Counting Collections would need to pay upfront costs of £688.12. The main expense is the Counting Library resources, which cost £551.93 at the time of writing. There are additional costs of £123 for the teacher professional development and £13.19 for access to the online learning environment during the training period. These are all one-off costs that do not need to be repeated each year. All teacher training is delivered online except for the first session, which is in person. Teachers should not need extra time to plan or prepare as the intervention is delivered to the whole class and takes place within existing maths teaching. Schools may need to provide cover for teachers attending the training sessions, which take place over one full day and four half-days. Paying for supply cover would amount to around £190 per full day based on the NASUWT Supply Teachers' Pay scale, totalling £570. This would mean that a school using supply teachers could expect to pay £1,258.12 to deliver Counting Collections. However, as these costs are approximate, they are not included in the calculations here. Once the teacher is trained and the materials are purchased, schools can continue running the intervention in subsequent school years at no additional cost.

This trial limited class sizes to 20 pupils per school, and this is the figure used to calculate per-pupil costs. Dividing the total cost by 20 pupils over three years (£688.12 ÷ 20 ÷ 3) gives a per-pupil cost of £11.47. All figures are presented below in Table 24. In Table 25, the cumulative costs are shown. As all costs are paid upfront, there are no increases to cumulative costs year on year.

Table 25: Cost of delivering Counting Collections

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
Counting Collections library resources	Start-up cost per school	£551.93	£551.93	
Teacher professional development	Start-up cost per school	£123.00	£123.00	
Online learning environment access	Start-up cost per school	£13.19	£13.19	
Total		£688.12	£688.12	e.g. £688.12 ÷ 3 ÷ 20 = £11.47

Table 26: Cumulative costs of Counting Collections assuming delivery over three years

	Year 1	Year 2	Year 3
Counting Collections	£688.12	£688.12	£688.12

Conclusion

Table 27: Key conclusions

Key conclusions	
1.	Children in Counting Collections schools made the equivalent of one month's progress in number attainment, on average, compared to children in the control schools. This result has a moderate to high security rating.
2.	Among children eligible for free school meals, those in Counting Collections schools made no additional months' progress in numeracy, on average, compared to children eligible for Free School Meals in other schools. These results may have lower security than the overall findings because of the smaller number of pupils.
3.	Teachers who had taken part in Counting Collections were more likely to report greater confidence and 'major improvements' in teaching and supporting pupils with early number interactions than teachers in control schools.
4.	Participating teachers reported that the programme improved pupil collaborative working, recording, mathematical thinking, and mathematical talk. They also reported increased engagement, concentration, motivation, number confidence, and enjoyment of mathematics.
5.	Observational findings indicated that most pupils were engaged in high quality number experiences supported by adult-child interactions, though some teachers found the four-part routine challenging and suggested more guidance for extending higher attainers and supporting lower attainers.

Impact evaluation and IPE integration

Evidence to support the logic model

The inputs specified in the logic model were observed during the trial, as evidenced through training observations, school fieldwork visits, and the teacher survey. Increased teacher confidence in planning and teaching whole-class maths sessions—following the Counting Collections four-part sequence—was evident and grew as the training and delivery progressed. Across the programme, children received the weekly Counting Collections four-part lesson.

In terms of intervention outputs, fieldwork observations and teacher interviews showed increased teacher confidence in their support of pupils and their early number skills in several areas improved teacher knowledge and confidence in planning activities teaching and interacting with pupils—particularly in building on pupil mathematical thinking and mathematical talk—and the use of learning trajectories to support progression were all evidenced. Teachers also felt more confident in seeing the learning potential through the pupils' mathematical jottings. As a result, pupils were more likely to engage in high quality number experiences in lessons and gain in number confidence. According to teachers, pupils enjoyed collaborating with a partner and using the materials from the Counting Collections library. The structure of the session facilitated peer-to-peer interactions and some of the number skills gained from the four-part sequence were transferred over into continuous provision. These findings align with survey data on teacher confidence in teaching early number, including the planning and assessment of early number, where respondents in intervention schools reported higher levels of confidence across several areas.

The outcomes evidenced in the teacher survey and school visits included improvements in pupil confidence and enjoyment in counting, for all children. There is also evidence from the impact evaluation of improving pupil attainment in maths, with the analysis finding that intervention pupils made the equivalent of one month of additional progress, although this was not the case for disadvantaged pupils. This may be associated with concerns expressed by teachers about differentiating for those insecure in mathematical concepts.

In the ten schools visited, practice was varied in terms of TA involvement in the programme so there is insufficient evidence to suggest that this aspect of the programme may have impact. Future iterations of the programme may consider training TAs to support delivery. It would also be worth exploring how the programme can be tailored to provide specific support for those pupils who are less secure in aspects of basic counting. This seems important given the greater progress made by non-disadvantaged pupils compared to those defined as disadvantaged even if the trial was not powered to detect effects of the sizes observed, and the known attainment gap between these groups.

Interpretation

The impact evaluation found that the intervention is associated with one month of additional progress (ES = 0.08) as measured by the primary outcome. Around one in four pupils that were randomised at the start of the trial did not complete outcome testing, which unfortunately weakens the evidence produced, however, having used multiple imputation to estimate outcome test scores from pupils that completed the baseline assessment and found a very similar effect size (0.09) to the headline ITT analysis, the findings appear to be robust.

Results from the analysis of disadvantaged pupils were somewhat different showed a stark contrast to those from the overall sample. There was no evidence that intervention FSM pupils made better progress than those in the control group: the evaluation has not discovered any evidence to explain why this is, although the trial was not powered to detect effect sizes of the magnitude observed in either the headline analysis or the subgroup of disadvantaged pupils. From the data collected, it is clear that disadvantaged pupils are starting school significantly behind their non-disadvantaged peers in number skills. It is also clear that this gap remains when pupils are at the end of their first year of compulsory schooling, as has been observed elsewhere with reference to data from the EYFSP (EPI, 2024). Counting Collections is a universal approach to developing number that is delivered to the whole class, which should benefit disadvantaged and non-disadvantaged pupils equally. It is possible that the lack of additional progress for FSM pupils is due to the disadvantage gap when they start school, but this is something to examine in future research.

The disadvantage gap may be reflected in the concerns expressed by teachers over ‘managing’ the learning outcomes for some lower attaining pupils, for example, in differentiating the number of objects in a collection, pupil counting strategies, and teachers’ concerns over not being able to offer one to one support where pupils demonstrate misconceptions. Letting such misconceptions go uncorrected created a tension for teachers. This may be related to a wider perspective on the incremental nature of mathematics learning held by many practicing teachers and the need to secure accurate counting and one to one correspondence and other fundamental mathematical concepts. TAs, when present, were mainly used to support the low attaining pupils and most had been advised on the delivery of the programme by the teacher. Significantly, observations showed that in some cases low attaining pupils sometimes struggled moving through the four-part routine and were not quite sure about how to strategise or record the collection. Schools approached this differently. Some focused on the counting skills and others gave continued support to navigate the different parts of the routine. Focused training for TAs in supporting the Counting Collections programme may be a positive way forward in developing the programme.

Limitations and lessons learned

The impact evaluation recruited enough schools for an MDES below 0.20 standard deviations. Repeating the power calculations using values obtained from the analysis sample suggested that the trial design was more sensitive than initially assumed. Compliance among intervention schools was high, and although compliance correlated with outcome test scores, the CACE analysis estimated very small differences in effect sizes. To account for the potential influence of other relevant interventions taking place alongside Counting Collections, the randomisation included a stratifier indicating whether each school was already implementing Mastering Number, which is used widely in England. This ensured that schools already using Mastering Number were distributed evenly between the two treatment groups. The impact evaluation found no evidence that Mastering Number had any effect on the results.

These positive factors are counteracted to some extent by trial attrition, which was 24%. This figure is an estimate based on a projected number of pupils per school as randomisation was conducted before schools supplied pupil lists to the evaluation team. Imbalance between the intervention and control groups was observed on the baseline measure, for which data was collected after schools were informed of their allocation. It is unlikely that this biased the baseline scores in itself as allocation was at school level and baseline testing was completed before intervention schools started to deliver the programme. However, eight of the 180 schools did not complete baseline testing despite having already been randomised, and control schools were more likely to withdraw at that stage than intervention schools.

Randomising schools before collecting pupil data or administering baseline assessments undoubtedly contributed to the attrition, although only around one in five pupils (n = 160) that did not complete outcome testing were lost from the trial at this point. Completed baseline test booklets from two schools were lost in the post and one school withdrew from the study before outcome testing. In total, these three schools accounted for only 60 pupils, less than 10% of overall attrition. Indeed, with outcome data collected in 169 of the 180 schools recruited, most attrition was due to individual pupils leaving school or being absent on assessment days. Mop-up visits achieved some success yet pupils with

persistent absence were still missed. While attrition appears to be the greatest threat to the validity of the results, missing data analysis and multiple imputation were both carried out and provide assurance that the main results are robust.

The lack of a baseline teacher survey is the main limitation of the IPE. One option would have been to run the survey just before randomisation when the pressure of baseline test data collection was not a constraint. It is possible that the difficulties of administering baseline assessments with pupils in their very first weeks of school to such a tight timeline were underestimated. The fact that so little evidence exists around pupil ability in number in the early part of the reception year attests to the challenges involved. Nevertheless, the post-intervention survey was designed to ask teachers to reflect on changes throughout the study year and achieved a good response rate (70%) in the final weeks of the trial.

Future research and publications

With the headline analysis for the whole sample showing that the intervention is associated with one month of additional progress, it is of some concern that this finding does not extend to disadvantaged pupils. Any gains made by disadvantaged pupils over the study period were eclipsed by those made by their non-disadvantaged classmates with the caveat that the trial was not powered to detect such effect sizes. The evaluation has not produced any evidence that might explain this, however, based on teacher interviews and observations, we have suggested why this may be the case and how it might be addressed in future iterations of the programme. One possibility is that a whole-class intervention such as Counting Collections is not geared toward particular groups of pupils. This is something to consider for future research, for example, an intervention that considers specific groups within the whole-class approach allowing for targeted adaptations through adult-child interactions. This would need to consider the resources made available to the early years setting and the business-as-usual model.

Having shown evidence of additional pupil progress associated with the intervention, the uncertainty about this result could be overcome by designing a trial with the power to detect a smaller effect size and reconsideration of the targeted professional development to include TAs, who are often the primary support for the most disadvantaged pupils. This would require a far higher number of schools to take part, which may be difficult to deliver in terms of recruitment, teacher training, and administering assessments. Organising a larger trial would require a high number of assessments during the early weeks of school, presenting obvious difficulties. One way in which these challenges could be mitigated is if the government would make the scores on the recently introduced Reception Baseline Assessment available through the NPD. At present there has been no announcement regarding a change of policy on this. Statutory assessments are taking place in schools without results being released for scrutiny, which places burden on schools without providing any information to advance knowledge about what works for teaching the youngest pupils in schools.

References

Bisaillon, N. (2023) 'Development of Number Sense and Numeration: A Continuum Hypothesis' ,*Journal of Research in Science, Mathematics and Technology Education*. 6, pp. 91–108.

Bloom, H. S. (1984) 'Accounting for No-Shows in Experimental Evaluation Designs'. *Evaluation Review*, 8 (2), pp. 225–246.

Bloom, H. S., Richburg-Hayes, L. and Black, A. R. (2007) 'Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions', *Educational Evaluation and Policy Analysis*, 29 (1), pp. 30–59.

Boylan, M. (2018) 'Where Did Maths Mastery Come from?', *Schools Week*. https://schoolsweek.co.uk/where-did-maths-mastery-come-from

Bruner, J. S. (1966) *Toward a Theory of Instruction*, Cambridge: Harvard University Press.

Chen, J., 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, pp. 367–377.

Coldwell, M. and Maxwell, B. (2018) 'Using Evidence-Informed Logic Models to Bridge Methods in Educational Evaluation' *Review of Education*, 6 (3), pp. 267–300.

Demack, S. (2019) 'Does the Classroom Level Matter in the Design of Educational Trials? A Theoretical and Empirical Review', London: Education Endowment Foundation.

Denton, K. and West, J. (2002) 'Children's Reading and Mathematics Achievement in Kindergarten and First Grade', Washington, DC: U.S. Department of Education, National Center for Statistics.

Dorwick, N. (2015) 'Numbers Count: A Large-Scale Intervention for Young Children Who Struggle with Mathematics', in R. C. Kadosh and A. Dowker (eds) *The Oxford Handbook of Numerical Cognition*, Oxford Library of Psychology (2015; online edn, Oxford Academic, 3 March 2014) (pp. 1099–117). https://doi.org/10.1093/oxfordhb/9780199642342.001.0001

Feikes, D. and Schwingendorf, K. (2008) 'The Importance of Compression in Children's Learning of Mathematics and Teacher's Learning to Teach Mathematics' *Mediterranean Journal for Research in Mathematics Education*, 7.

Göbel, S. M., Watson, S. E., Lervåg, A. and Hulme, C. (2014) 'Children's Arithmetic Development: It is Number Knowledge, Not the Approximate Number Sense, that Counts' *Psychological Science*, 25 (3), pp. 789–798.

Gripton, C. and Pawluch, D. (2021) 'Counting Collections in the Early Years', *Mathematics Teaching*, 275, pp. 6–10. https://atm.org.uk/write/MediaUploads/Journals/MT275/02.pdf

Henry, G., Mashbur, A. and Konold, T. (2007) 'Developing and Evaluating a Measure of Young Children's Attitudes Towards School and Learning', *Psychoeducational Assessment*, 25 (3), pp. 271–284.

Jordan, N. C., Glutting, J. and Ramineni, C. (2010) 'The Importance of Number Sense to Mathematics Achievement in First and Third Grades', *Learning and Individual Differences*, 20 (2), pp. 82–88.

Manfra, L., Dinehart, L. and Sembiante, S. (2014) 'Associations Between Counting Ability in Preschool and Mathematic Performance in First Grade Among a Sample of Ethnically Diverse, Low-Income Children' *Research in Childhood Education International*, 28, pp. 101–114.

Miles, M. B., Huberman, A. M. and Saldana, J. (2019) *Qualitative Data Analysis: A Methods Source Book* (4th edn), London: Sage.

NCETM (2024) 'Mastery Myth-Busting', National Centre for Excellence in the Teaching of Mathematics. https://www.ncetm.org.uk/features/mastery-myth-busting/

Ofsted (2021) 'Research Review Series: Mathematics'. https://www.gov.uk/government/publications/research-review-series-mathematics/research-review-series-mathematics

Ofsted (2022) 'Early Years Research Review: Best Start in Life, Part 1'. https://www.gov.uk/government/publications/best-start-in-life-a-research-review-for-early-years/best-start-in-life-part-1-setting-the-scene#teaching

Perry, B., Gervasoni, A., Hampshire, A., and O'Neill, W. (2016) 'Let's Count: Improving Community Approaches to Early Years Mathematics Learning, Teaching and Dispositions through Noticing, Exploring and Talking about Mathematics', in B. White, M. Chinnappan, and S. Trenholm (eds) *Opening up Mathematics Education*

Research: Proceedings of the 39th annual conference of the Mathematics Education Research Group of Australasia, Mathematics Education Research Group of Australasia (pp. 75–82).

Raghubar, K. P. and Barnes, M. A. (2017) 'Early Numeracy Skills in Preschool-Aged Children: A Review of Neurocognitive Findings and Implications for Assessment and Intervention', *The Clinical Neuropsychologist*, 31 (2), pp. 329–351.

Sasanguie, D., van den Bussche, E. and Reynvoet, B. (2012) 'Predictors for Mathematics Achievement? Evidence From a Longitudinal Study', *Mind, Brain, and Education*, 6, pp. 119–128.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J. and De Smedt, B. (2017) 'Associations of Non-Symbolic and Symbolic Numerical Magnitude Processing with Mathematical Competence: A Meta-Analysis', *Developmental Science*, 20 (3), pp. 1–16.

Short, D. S. and McLean, J. F. (2023) 'The Relationship Between Numerical Mapping Abilities, Maths Achievement and Socioeconomic Status in 4- and 5-Year-Old Children', *Educational Psychology*, 93, pp. 641–657.

Shulman, L. S. (1986) 'Those Who Understand: Knowledge Growth in Teaching', *Educational Researcher*, 15 (2), pp. 4–14.

Torgerson, C., Wiggins, A., Torgerson, D. J., Ainsworth, H., Barmby, P., Hewitt, C., Jones, K., Hendry, V., Askew, M., Bland, M., Coe, R., Higgins, S., Hodgen, J., Hulme, C. and Tymms, P. (2011) '*Every Child Counts: The Independent Evaluation – Technical Report.* Department for Education. Ref: DFE-RR091A. Available at: https://www.gov.uk/government/publications/every-child-counts-the-independent-evaluation-technical-report

Turvill, R. (2021) 'Making it Count: Thriving with Maths Schemes in Year 1', Association of Teachers of Mathematics. https://atm.org.uk/write/MediaUploads/Journals/MT276/07.pdf

Appendix A: EEF cost rating

Figure 2: Cost Rating

Cost rating	Description
£ £ £ £ £	*Very low*: less than £80 per pupil per year.
£ £ £ £ £	*Low*: up to about £200 per pupil per year.
£ £ £ £ £	*Moderate*: up to about £700 per pupil per year.
£ £ £ £ £	*High*: up to £1,200 per pupil per year.
£ £ £ £ £	*Very high*: over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

And Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**			Adjustment for threats to internal validity **[0]**		
5	Randomised design	<= 0.2	0–10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11–20%					
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21–30%	3				3
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31–40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41–50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Moderate	There is imbalance in baseline attainment (0.09 SD), however it sounds like the headline analysis included baseline attainment as a covariate, and a sensitivity check was run without covariates. To check this completely I would need to see the appendices with effect sizes across models. **As the control schools had higher baseline scores, I would expect this to boost their endpoint scores, deflating the effect size.**
Threat 2: Concurrent Interventions	Low	Mastering Number is the most prevalent/relevant intervention, this was used as a stratification variable and there was good balance at baseline. (Low risk) The IPE did not report on additional interventions used in control schools (e.g. via endpoint survey). School visits (n=10) suggest there was a substantial use of mastery-based teaching schemes in intervention classes, these schemes target similar pupil outcomes.
Threat 3: Experimental effects	Low	Well designed and conducted
Threat 4: Implementation fidelity	Low	Compliance was aligned with the logic model and was very high.
Threat 5: Missing Data	Moderate	The total missing data is 24% - estimated maximum, as total number of pupils was not known at randomisation. Based on this estimate, 16% of the sample was lost before/at baseline, while the remaining 8% was lost at follow-up. While missing data analysis has been carried out and suggests no substantial difference, this has only been done for pupils with baseline data, which accounts for less than half of the missing data. There was also differential attrition pre-baseline which

		threatens validity, with higher attrition in the control group. **Higher attrition in the control group most commonly inflates the effect size.**
Threat 6: Measurement of Outcomes	Low	SENT-R has high relevance and construct validity, test administrators were blinded to allocation, and no ceiling/floor effects are shown in the score distributions.
Threat 7: Selective reporting	Low	Study is pre-registered and impact evaluation follows the published study plan and SAP.

Initial padlocks: 5 Padlocks

Reason for adjustment for threats to validity: 2 Padlocks removed – see above

Final padlock score: initial score adjusted for threats to validity = 3 Padlocks

Appendix C: Effect size estimation

Appendix Table 2: Effect size estimation

			Intervention group		Control group			
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance	Population variance (if applicable)
SENT-R(B)	26.48-26.29 =0.19	0.77/9.65 =0.08	1412(308)	93.64	1329(291)	92.81	93.21	

Figure 2:Baseline scores

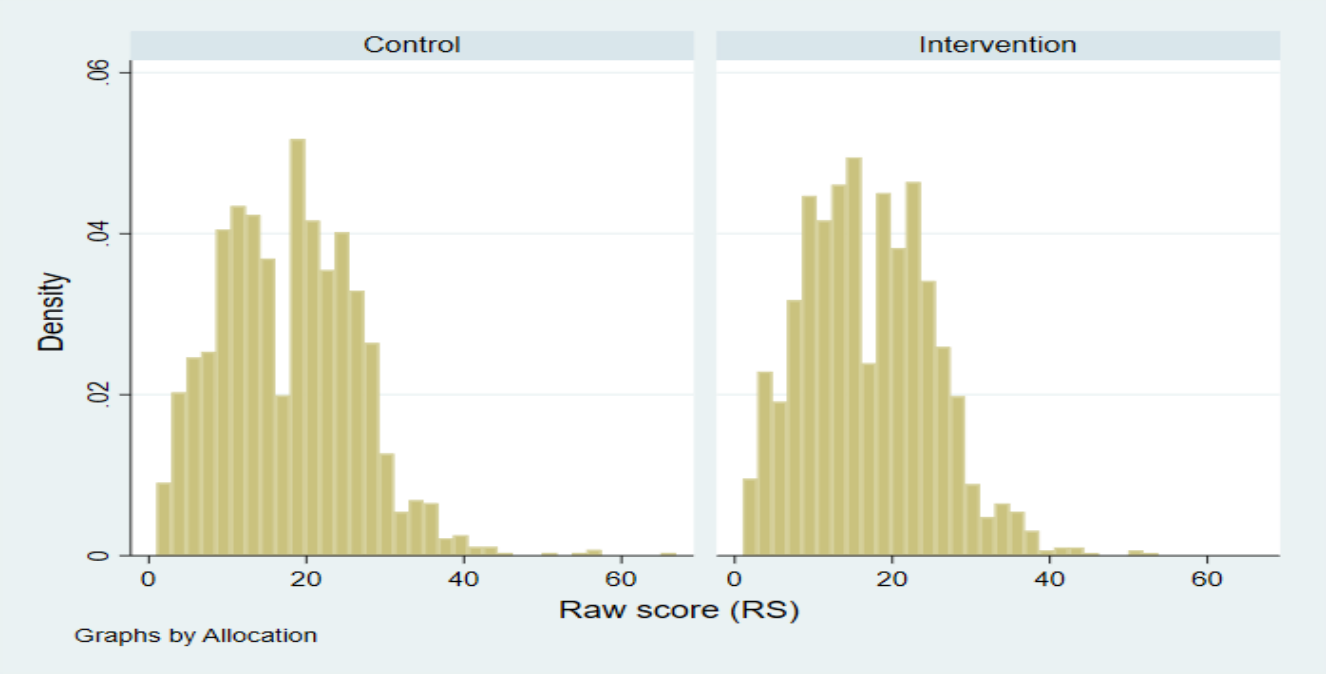
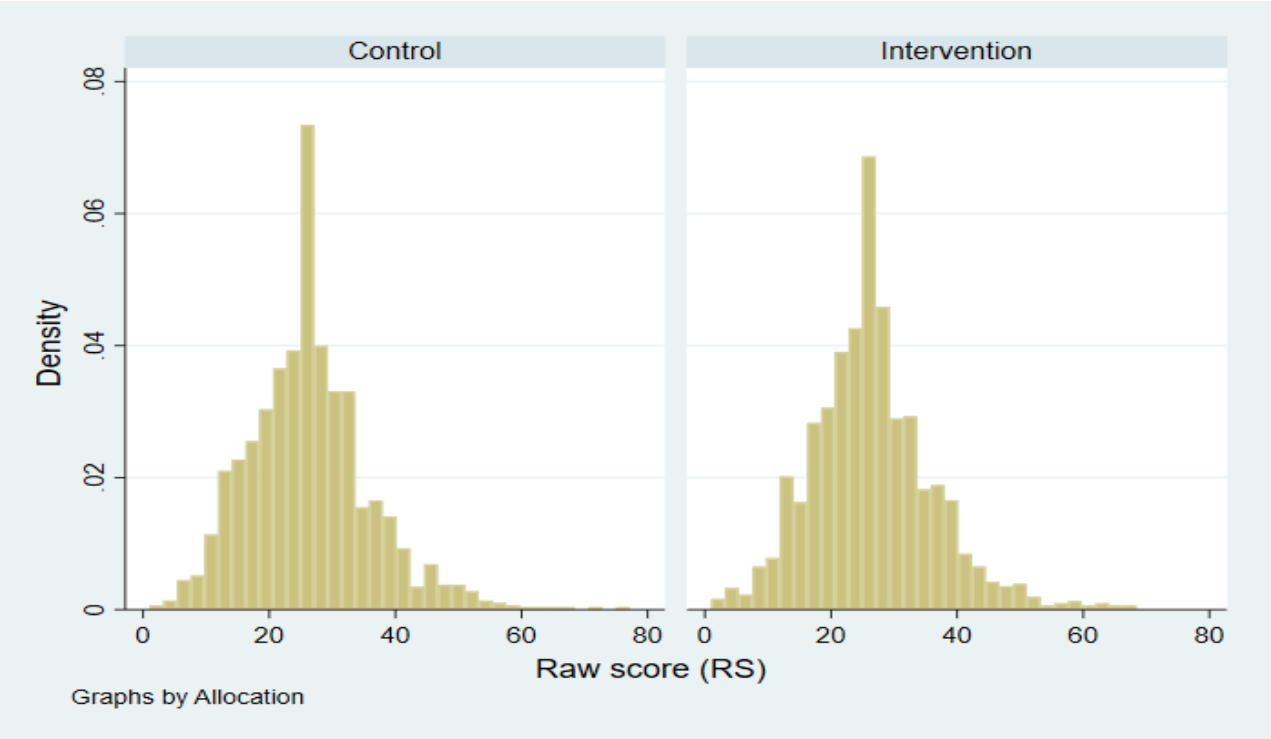


Figure 3: Outcome scores



Further appendices:

Please find the further appendices as a separate document on the project page.

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

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

https://educationendowmentfoundation.org.uk
@EducEndowFoundn
Facebook.com/EducEndowFoundn