Independent evaluation of Maths Champions in nursery to develop children's early numeracy: A two-armed cluster randomised controlled trial

Evaluation Report

July 2023

Updated December 2023

Second update July 2024

Lyn Robinson-Smith, Katie Whiteside, Carole Torgerson, Caroline Fairhurst, Danielle Podmore, Tom Davill, Kerry Bell, Xiaofei Qi, Heather Leggett, Louise Elliott, Catherine Hewitt, Kalpita Baird, Victoria Menzies, David Torgerson, Hannah Ainsworth

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 – 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** 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

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk


Department
for Education


What
Works
Network

Contents

About the evaluator

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

Lyn Robinson-Smith, Katie Whiteside, Carole Torgerson, Caroline Fairhurst, Danielle Podmore, Tom Davill, Kerry Bell, Xiaofei Qi, Heather Leggett, Louise Elliott, Catherine Hewitt, Kalpita Baird, Victoria Menzies, David Torgerson, Hannah Ainsworth. The co-principal investigators were Dr Lyn Robinson-Smith, Hannah Ainsworth (until August 2022), and Professor David Torgerson (from August 2022).

Contact details

Dr Lyn Robinson-Smith, York Trials Unit, Department of Health Sciences, University of York

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

Acknowledgements

The evaluation team would like to thank all the nurseries, practitioners, teachers, parents and carers, and children that were involved in this evaluation for giving their time and efforts in participating in the evaluation of Maths Champions.

We would also like to thank the team at the National Day Nursery Association for their cooperation in all aspects of this evaluation; it has been a pleasure to work with you all.

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

Revisions to report

This report was first published in July 2023.

The report was revised in December 2023 with an updated cost evaluation which include more accurate estimates of the cost of the programme to settings as delivered in this trial. Whilst the cost per setting and child have both been revised, both the original and revised analysis categorise this programme as a very low cost programme to implement, and do not impact the conclusions on the efficacy and cost effectiveness of the programme.

The report was revised again in June 2024 to add in results of a pre-planned longitudinal analysis, looking at long-term outcomes collected from the National Pupil Database (NPD) for children recruited into the trial. A new conclusion was added to the executive summary to describe the findings, and the full additional exploratory analysis was added as an appendix to the report (Appendix F).

Revised executive summary

The project

The Maths Champions programme builds the knowledge of nursery practitioners to support children's early mathematical development. It is a one-year programme developed and delivered by the National Day Nurseries Association (NDNA), to both private, voluntary, or independent (PVI) and school-based nursery (SN) settings. In each setting, a 'Maths Champion' (MC) is selected—a senior member of staff (either a graduate or level 3 practitioner) who undertakes the training and is responsible for developing the maths teaching of the other practitioners working with children in their setting. A deputy MC (DMC) is also nominated by the nursery to support the MC with the delivery of the programme (who is also qualified to at least Level 3). The programme consists of online training covering topics such as early years maths theory and how to support other staff in the nursery, an audit tool to evaluate maths provision in the nursery, support to develop and review an action plan to improve maths provision, support tracking progress of children, access to online resources, and tailored one to one remote support.

One hundred and thirty-four nurseries and 1,304 children took part in this effectiveness trial, which was a two-armed randomised controlled trial with randomisation at the nursery level. The trial evaluated the impact of the Maths Champions programme on the maths attainment of children aged three to four after approximately seven months of intervention delivery using the Assessment Profile on Entry for Children and Toddlers (ASPECTS) maths score. Secondary outcome measures included a measure of language (reading and phonological awareness) and practitioner confidence in teaching children maths. Longer term outcomes will be measured using participating children's Early Years Foundation Stage Profile data (completed at the end of reception 2023) and will be published in an addendum report in spring 2024. An implementation and process evaluation (IPE) comprised surveys from all nurseries, interviews with intervention nurseries, and a sample of nurseries conducted a measure of the quality of the nursery environment.

A pilot study was conducted prior to the effectiveness trial to pilot a number of changes to the MC programme before implementing these in the effectiveness trial, and to pilot elements of the evaluation methodology. The effectiveness trial intervention delivery period ran from October 2021 to June 2022 (a year later than originally planned due to COVID-19 setting closures). This project was funded by the EEF and the Department for Education.

Table 1: Key conclusions

Key conclusions	
1.	Children in nurseries allocated to the intervention group made, on average, the equivalent of three months' additional progress in maths attainment compared to children in control nurseries. This result has a very high security rating.
2.	Children in nurseries allocated to the intervention group made, on average, the equivalent of three months' additional progress in language attainment compared to children in control nurseries.
3.	Children eligible for Early Years Pupil Premium (EYPP) in the intervention nurseries made, on average, the equivalent of six months' additional progress in maths attainment compared to children eligible for EYPP in control nurseries. These results, while promising, should be treated with more caution than the analysis on all pupils as fewer children were included in this analysis.
4.	The results from the implementation and process evaluation support the majority of the components of the logic model, specifically the training and support, the MC and DMC roles, and the action plans, suggesting these should be maintained in the future.
5.	The implementation and process evaluation found that the commitment of the MC and DMC is crucial to successful implementation and that the roles can be interpreted flexibly depending on the needs of the nursery and the choice of nominated MC and DMC.
6.	Subsequent analysis was conducted using Early Years Foundation Stage Profile (EYFSP) assessment data from the National Pupil Database (NPD), to assess the potential long-term impact of Maths Champions on children's attainment. This exploratory analysis found that children in nurseries allocated to the intervention group made, on average, the equivalent of two months' additional progress in mathematics and language attainment compared to children in control nurseries, as well as being slightly more likely to attain a Good Level of Development (GLD) compared to children in control nurseries (equating to approximately one months' additional progress). Children eligible for EYPP in the intervention nurseries made, on average, the equivalent of six months' additional progress in mathematics attainment compared to children eligible for EYPP in control nurseries (however these results, while promising, should be treated with more caution than the analysis on all pupils as fewer children were included in this analysis). With all these estimates there is some statistical uncertainty, as the impact estimate ranges

are wide, and therefore should be treated with caution. Full findings from this exploratory longitudinal analysis can be found in Appendix F[1].

EEF security rating

These findings have a very high security rating. This was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of settings. The trial was a well-designed two-armed randomised controlled trial and was well-powered. Relatively few children (7%) who started the trial were not included the final analysis.

Additional findings

Children attending nurseries that were allocated to receive the Maths Champions programme made, on average, three additional months' progress in maths, as measured by ASPECTS, than those in the control group equivalent. In addition, children in MC nurseries were estimated to have made the equivalent of three additional months' progress in the secondary outcome measure of language, also measured using ASPECTS. It was intended that increased use of mathematical language would, in turn, expand and improve children's use of language more generally. This is our best estimate of impact, which has a very high security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme ranges from two to five months of additional progress for maths and one to four months for language for children in the intervention arm. Children attending nurseries receiving the MC programme who were eligible for the Early Years Pupil Premium (EYPP) made, on average, the equivalent of six months' additional progress in maths compared to EYPP-eligible children in control nurseries (with the possible effect ranging from three to nine months). This result, while promising, should be treated with more caution than the primary result as fewer children were included in this analysis.

There was some evidence that practitioners' confidence in helping nursery-aged children learn maths was improved for MCs and DMCs in intervention nurseries, which was an important hypothesised route to impact. The effect of the intervention on maths was not significantly different between boys and girls, nor did it change substantially depending on the average number of hours the child attended the nursery per week, eligibility for Free Early Education Entitlement (FEEE), or type of nursery (PVI or SN). The majority view was that there were no barriers to implementing the MC programme in nurseries. The IPE analysis results are concordant with the results from the impact evaluation, with staff sharing their experiences of the noticeable improvement in children's progress. The IPE indicated that the DMC role enabled the MC to effectively share responsibility of implementing the MC programme. The findings of both the impact evaluation and the IPE support the majority of the components of the logic model.

This is the second effectiveness trial of the Maths Champions programme funded by the EEF. The first trial found that children in the intervention group made, on average, the equivalent of two additional months' progress in maths and language in comparison to the control group, although the study suffered high attrition (36% of the children who were recruited to the trial did not complete the post-test) and therefore had a low EEF security rating (two padlocks) meaning we could not be confident in the results. In this re-trial at effectiveness level, the evaluation was designed to minimise attrition, and did so successfully. The very high security rating of this trial gives high confidence that the programme has a positive impact on children's maths attainment. The impact results of this trial align to the evidence within the EEF's Teaching and Learning Toolkit, which reports high quality CPD is effective at improving child attainment within the early years and that supporting practitioner knowledge of maths and children's mathematical development supports early mathematical learning. This trial has shown the MC programme to be a cost effective and scalable way to improve child outcomes in PVI nurseries and SN in England.

[1] This analysis was conducted later than the publication of the original report in December 2023, and therefore this conclusion was added in June 2024 (full analysis can be found in Appendix F).

Cost

The estimated cost of Maths Champions is £2,896.80 per nursery over a three-year period, or £28.39 per child per year when averaged over three years[2]. This figure is based on 34 children per nursery per year and includes costs for staff training, optional cover, and optional materials.

Impact

Table 2: Summary of impact on primary outcome

Outcome/ group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No of children (intervention; control)	P Value	EEF cost rating
Mathematics (ASPECTS)	0.25 (0.12 to 0.38)	3	🔒🔒🔒🔒🔒	1,209 (600; 609)	0.001	£ £ £ £ £
Mathematics (ASPECTS) (EYPP subgroup)	0.47 (0.20 to 0.73)	6	N/A	154 (66; 88)	0.001	N/A

[2] The original version of this report, published in July 2023, contained inaccurate cost figures. An update in December 2023 rectifies those inaccuracies and includes a revised cost analysis.

Introduction

Study rationale and background

In England, the most recent Key Stage 2 assessments show that just 71% of children meet the expected standard in maths at the end of primary school (DfE, 2022), a reduction from 79% in 2019 (DfE, 2019c). To help minimise attainment gaps, it is important to support children's early maths development as maths skills at school entry are predictive of both later maths attainment and general educational attainment (Duncan et al., 2007; Watts et al., 2014). Quality preschool provision, with an enriching numeracy curriculum, is important to support children's maths development and long term outcomes (Asmussen et al., 2018). Preschool attendance and quality of preschool provision, as well as preschool effectiveness in promoting early number concepts, are predictive of children's maths and reading attainment at Key Stages 1 and 2 (Sammons et al., 2004; 2008), maths and science attainment at Key Stage 3 (Sammons et al., 2011), and even GCSE results (Sylva et al., 2014). Despite this, many nursery practitioners lack training in maths provision and do not feel confident in their own maths skills (von Spreckelsen et al., 2019).

The National Day Nurseries Association (NDNA) developed the Maths Champions (MC) programme with the aim of building the confidence and knowledge of nursery practitioners to support the development of children's early maths skills. The EEF have previously commissioned an effectiveness trial evaluating the NDNA's MC programme delivered to graduate practitioners in private, voluntary, and independent (PVI) nurseries during 2016/2017 (Robinson-Smith et al., 2018). Here, children aged three to four from nurseries randomly allocated to use the MC programme made the equivalent of two months' additional progress in maths and language development (reading and phonological awareness) in comparison to children in the control group. This effect on maths development was not affected by children's eligibility for Early Years Pupil Premium, how many hours a child attended nursery, or gender. However, the findings should be viewed with caution as the results were not statistically significant. Moreover, while the trial was well designed and conducted, it suffered from high attrition. Indeed, 36% of children recruited into the trial were not included in the primary analyses; more than half of this attrition was attributable to children who were assessed at pre-test leaving the nursery prior to post-testing. The level of attrition was a potential threat to the validity of the study's findings. Consequently, the EEF funded a second evaluation of the NDNA's MC programme to be implemented in two phases: (1) the pilot study and (2) the effectiveness trial. The pilot study was a non-randomised feasibility study predominately designed to pilot a number of changes to the MC programme before implementing these in the second effectiveness trial and, to a lesser degree, to pilot elements of the evaluation methodology. An interim report with findings from the pilot was written prior to starting the effectiveness trial; a summary of the pilot methodology and findings is presented in Appendix E. This report focuses on the effectiveness trial.

The first effectiveness trial of MC (Robinson-Smith et al., 2018) found that the programme was positively received by many nurseries, with 82% of nurseries being at least minimally engaged with the intervention, defined as the nursery being rated as being very or partially engaged in at least one of the eight pre-specified core aspects of the intervention; however, some nurseries raised the burden on staff time as a significant issue. Staff often reported that they had to complete activities for the MC programme within their own non-working time, particularly during the set-up phase, as nurseries lacked the financial resources to free staff from their normal day to day work commitments. Nurseries that shared the workload of the programme amongst staff, rather than just being the responsibility of the Maths Champion (MC), exhibited higher levels of engagement. A core component of the programme at the time was the completion of the Basic and Key Skill Builder (BKSB), which required practitioners to complete an assessment at the start and end of the trial to assess maths skills as well as relevant online modules in between. Only 52% of nurseries engaged with this core component and practitioner interviews indicated that this was a significant barrier to engagement, negatively impacting on staff confidence and, as a consequence, becoming a barrier to engagement and implementation.

In response to the results of the first effectiveness trial of the MC programme, the delivery team made a number of changes to the programme (summarised in Appendix C). The most significant were the exclusion of the BKSB, the introduction of a Deputy MC (DMC; this role has a dual function allowing nurseries to spread the programme's workload and to enable continuity with the programme should the MC be absent or leave the nursery), and a move from face to face initial training to online training. All these changes essentially aimed to address the issue of staff burden reported by Robinson-Smith et al. (2018).

The removal of the BKSB came about due to low engagement with this component of the programme as reported in the first MC trial. As a result, the delivery team adapted the programme to provide practitioners with a more comprehensive

understanding of the main areas of early years maths (the following information about the programme provides further detail or correction to what is in the published trial protocol). Webinars and online modules were updated to help both the Champions and practitioners to gain a deeper understanding of the six main areas that collectively underpin children's early mathematical learning and provide the firm foundations for the maths that children will encounter as they go up the years in primary school. Between March 2020 and the start of the second trial, MC II, in September 2021 there was a change to the Early Years Foundation Stage framework (EYFS), which influenced the areas of maths learning in the programme. These six main areas include:

1. Cardinality and counting

   The cardinal value of numbers is important, so children know what the numbers mean in terms of how many things they refer to. Counting is one way to establish how many things are in a group. This area included subitising—recognising numbers without counting, which is an effective way for children to gain number meaning. More was added about this as a new aspect of number that practitioners were not so aware of.

2. Comparison

   Comparing numbers involves knowing which numbers are worth more or less than each other. However, this depends both on understanding cardinal values and numbers. This understanding helps underpin the mental number line which children will develop.

3. Composition

   Knowing numbers are made up of two or more other smaller numbers. Learning to see a whole number and its parts at the same time is key to development in children's number understanding. Partitioning numbers into other numbers and putting them back together again underpins understanding of addition and subtraction.

4. Shape and spatial awareness

   Understanding shape properties and being able to visualise spatially in order to develop wider mathematical thinking and problem-solving as recent research highlights. More was added about spatial awareness and reasoning to reflect its inclusion in the EYFS educational programme.

5. Pattern

   Seeking and exploring patterns is at the heart of maths. Developing an awareness of pattern helps young children to notice and understand mathematical relationships and this can provide foundations of algebraic thinking. This was also added to reflect its emphasis in the EYFS educational programme.

6. Measure

   Evaluated in the first MC effectiveness trial, this provided firm foundations for children's development in this area; however, improvements were made to increase practitioner understanding. Recognising attributes and comparison are prerequisites for understanding the use of units for older children.

The Characteristics of Effective Teaching and Learning and sustained shared thinking were also emphasised as important aspects of early years maths pedagogy.

Changes that were made to the evaluation design between the previous and current effectiveness trials are also summarised in Appendix C. One change related to who within the nursery could be trained to be the MC. In the first effectiveness trial (Robinson-Smith et al., 2018), the inclusion criteria required nurseries to have at least one graduate (degree level) practitioner within the nursery who would be the nominated MC. Within this trial, the practitioner qualification requirements were lowered so nurseries without a graduate practitioner could also participate. In this trial, practitioners qualified to at least level 3 (A-level/NVQ level 3 or equivalent) who were responsible for leading the quality of the Early Years Foundation Stage (EYFS) at their nursery could receive training to become a MC. This was reflective of the changing landscape of early years practitioner qualifications. The NDNA's (2019) annual workforce survey demonstrated a reduction in the proportion of nursery staff with graduate qualifications in recent years. The Department for Education (2019b) reported that only 47% of private or voluntary nurseries had a graduate practitioner in 2019. Furthermore, while only 7% of staff in private or voluntary nurseries were graduate practitioners, 65% were qualified to level 3 (DfE, 2019b).

Another difference between this trial and the trial reported by Robinson-Smith et al. (2018) is that both PVI and school-based nurseries (SNs—including maintained nurseries) were recruited. This reflects national provision as, excluding childminders, SNs make up 27% of early years providers in England (DfE, 2019a). This second effectiveness trial sought to understand whether the MC programme could also be effective in SNs. To enable the inclusion of SNs, the intervention period was intended to be slightly longer in this trial compared to the previous trial, with nurseries receiving approximately seven or eight months of the intervention before outcome post-tests were conducted, rather than six to seven months in the previous trial.

A final point to note, in reference to policy changes within early years childcare, is that there has been a significant change to the government Free Early Education Entitlement (FEEE) scheme since Robinson-Smith et al. (2018) conducted their study. Since September 2017, FEEE has extended funded childcare from 15 to 30 hours per week (term time only) for all eligible three- to four-year-olds. In line with this policy change, we postulated at the design phase of the trial that children's average weekly attendance at nursery may increase (children attended nursery approximately 24 hours per week during the first MC trial) and, if so, children would have greater exposure to the intervention.

The trial will also assess the feasibility of accessing Ages and Stages Questionnaire (ASQ-3) data from NHS Digital—completed when participating children were two years old—to gauge if this data correlates to maths and language development at three and four years old.

Intervention

A description of the NDNA's MC programme is provided in Robinson-Smith et al. (2022, p.15) and cited here for ease of reference.

Description of the MC programme using the Template for Intervention Description and Replication (TIDieR)

Brief name

Maths Champions (MC).

Why—rationale, theory or goal of the elements essential to the programme

MC was developed in response to a number of challenges identified in the early years:

- there is an attainment gap in EYFSP results between disadvantaged children and their peers (Asmussen et al., 2018);
- early years practitioners have low confidence and professional understanding to support children's mathematical learning (All Party Parliamentary Group for Maths and Numeracy, 2014); and
- research tells us that children who start behind, stay behind (Asmussen et al., 2018).

The Early Intervention Foundation (Asmussen et al., 2018, p.149) concludes that enriching the maths curriculum in preschool results in gains for low-income children; Frye et al. (2013), from the What Works Clearinghouse, recommend embedding maths in daily routines and activities and using learning trajectories to monitor progress.

The goals of the MC programme are to:

- reduce the attainment gap in EYFSP results between disadvantaged children and their non-disadvantaged peers;
- increase early years practitioners' confidence and professional understanding to support children's mathematical learning; and
- provide children with the best start in mathematical development.

In line with recommendations by the EEF (2020) to improve maths in the early years, the MC programme was developed to achieve these goals by:

- increasing practitioners' understanding of how children learn maths;

- increasing understanding of pedagogy—for example, embedding maths though the day through direct teaching and sustained shared thinking;
- champions auditing practice and practitioners' knowledge and developing and reviewing plans of action with the aim of increasing children's achievement;
- champions working with practitioners in nurseries to develop mathematical understanding, skills, and confidence; and
- providing tools and resources to put learning into practice within their nurseries.

Who—recipients

Recipients of the programme were:

- private, voluntary, and independent nurseries, maintained nurseries, and school-based nurseries;
- a graduate or level 3 practitioner received training and support for the role of MC;
- another practitioner at each nursery, typically a room leader, qualified to at least level 3, received training and support for the role of deputy MC (DMC);
- the MC and DMC supported other practitioners in their nurseries to develop their professional understanding and confidence through, for example, coaching to improve practice;
- the MC programme was developed to work to improve maths provision and attainment in maths for all children in participating nurseries; and
- nurseries could encourage parents and carers regarding their child's mathematical development using face to face communication or by sharing resources from children's learning journals.

What—physical or informational materials

MCs and DMCs were provided with:

- an online webinar induction (one hour in duration) covering information about leading the programme in their nursery;
- information about the audit tools that MCs used to evaluate early years maths teaching in their nursery, for example, the quality of resources available in the nursery to support mathematical learning, staff use of mathematical language in discussions with children, and planning opportunities for mathematical learning in play and activities;
- access to three online courses (each approximately two hours in duration) made up of e-learning modules covering early years maths theory and how to support other staff in the nursery including—
    - coaching as an educational lead mathematical concepts in early years;
    - developing mathematical confidence in the early years: the big ideas of number sense; and
    - developing mathematical thinking in the early years: shape space, measures, and pattern, including characteristics of effective learning and sustained, shared thinking;
- access to an online platform with over 700 resources including number songs and rhymes, outdoor maths ideas, and links to useful websites and research (requirement to use ten mandatory resources from the platform, details below); and
- access to optional monthly webinars—developed in response to nursery's action plan themes, for example, using outdoor play and snack time to develop children's maths and the confidence of staff.

What—procedures, activities, and processes

Programme delivery involved the following:

- the use of an audit tool comprising of a set of survey questions, the development of an action plan, a review of the action plan with NDNA throughout, and a review of the action plan at the end of the support provided by NDNA;
- the opportunity to (optionally) track and monitor children's development, in line with the EYFS, for six children on a termly basis:[3] the use of the NDNA tracking tool was optional as nurseries may maintain their own tracking systems;
- the use of ten mandatory resources, provided through an online platform, for three- to four-year-olds: 'build a maze', 'number hunt', 'delivering the post', 'mud kitchen', 'cars down a ramp', 'patterns', 'construction', 'tidy up time', 'snack time', and 'outdoor games';
- the use of ten mandatory resources, provided through an online platform, for two- to three-year-olds: 'block play', 'tidying up', 'parachute games', 'number rhymes', 'snack time', 'small world', 'puzzles and shape sorters', 'let's picnic', 'sand and water play', and 'care routines';
- the MC lead at the NDNA provided one to one support to nurseries monthly, via telephone or video conference, to keep nurseries on track with the programme; and
- a case study—a portfolio review—completed by the NDNA: this included all steps nurseries had to undertake to be compliant with the programme, with particular regard to the nursery audit and following changes to tracked children's development (optional component).

Who—programme providers and implementers

The NDNA provided the programme and provided MCs and DMCs with training and support.

The MC (qualified in childcare to at least level 3) ran the programme within their nursery. With support from the NDNA, their responsibilities included the completion of online training, completing audits of maths teaching in their nursery, creating action plans for improving maths provision across the nursery, and working with other nursery staff to improve their practice and confidence in maths.

The DMC (typically a room leader at level 3) supported the MC to implement change and observe and track children. It was intended that the DMC may replace the MC if the MC were to leave the nursery. Alternatively, a new staff member may be trained up to replace the MC. The role of the DMC may be replaced, as appropriate, should the existing DMC take over the MC role or leave.

It was intended that practitioners within the nursery, with support from the MC and DMC, would implement change and observe and track children with increasing confidence.

How—mode of delivery

Training for MCs and DMCs was delivered though online webinars and e-learning modules. The NDNA additionally provided MCs with one to one support, mainly through monthly phone calls or video calls, if they needed additional support.

Where—location of delivery

The programme is available nationally (for the purpose of this trial, recruitment was initially geographically restricted to East and West Midlands, and then extended to other areas to meet recruitment targets).

[3] NDNA guidance for this step: 'The children you track should be carefully selected in order to show the best possible impact. We recommend selecting a range of children according to the composition of your nursery. This could include; a mixture of boys and girls, children with SEND, children who attend AM or PM only, children who attend 15 hours or 30 hours etc.'

As noted above, training for MCs and DMCs was online, and support was remote.

When and how much—duration and dosage

Usually, within the MC programme, nurseries are supported for a minimum of 12 months; in the context of this second effectiveness trial, however, nurseries were supported for seven to eight months (see the Study rationale and background section).

Tailoring—adaptation of the programme

The audit was conducted to identify needs and the action plan that MCs put together to improve maths provision was nursery specific.

The NDNA provided tailored one to one support to MCs throughout the year and particularly when putting together their action plans. Support was remote (mainly monthly phone calls). MCs could request additional support phone calls or web calls if necessary. Attendance at the optional webinars was dependent on audits and action plans.

How well (planned)—strategies to maximise effective implementation

In addition to the extensive training, resources, and support outlined in the sections above, the following strategies were employed to maximise effective implementation:

- the induction, used to gain commitment to programme, took MCs and DMCs step-by-step through the process and familiarised them with the early years development zone (online platform);
- a handbook was provided, which contained an overview of the programme steps and delivery schedule along with instructions for enrolling and important contact details for staying connected throughout their journey;
- phone calls were used rather than email; web calls were also used; and
- the potential for face to face visits at nursery to evaluate how revised resources were working in practice (not support visits); photographic evidence of the programme being implemented was collected for future programmes;

There was no requirement for additional resources as everything needed to implement the programme would be already available within the nursery.

Logic model

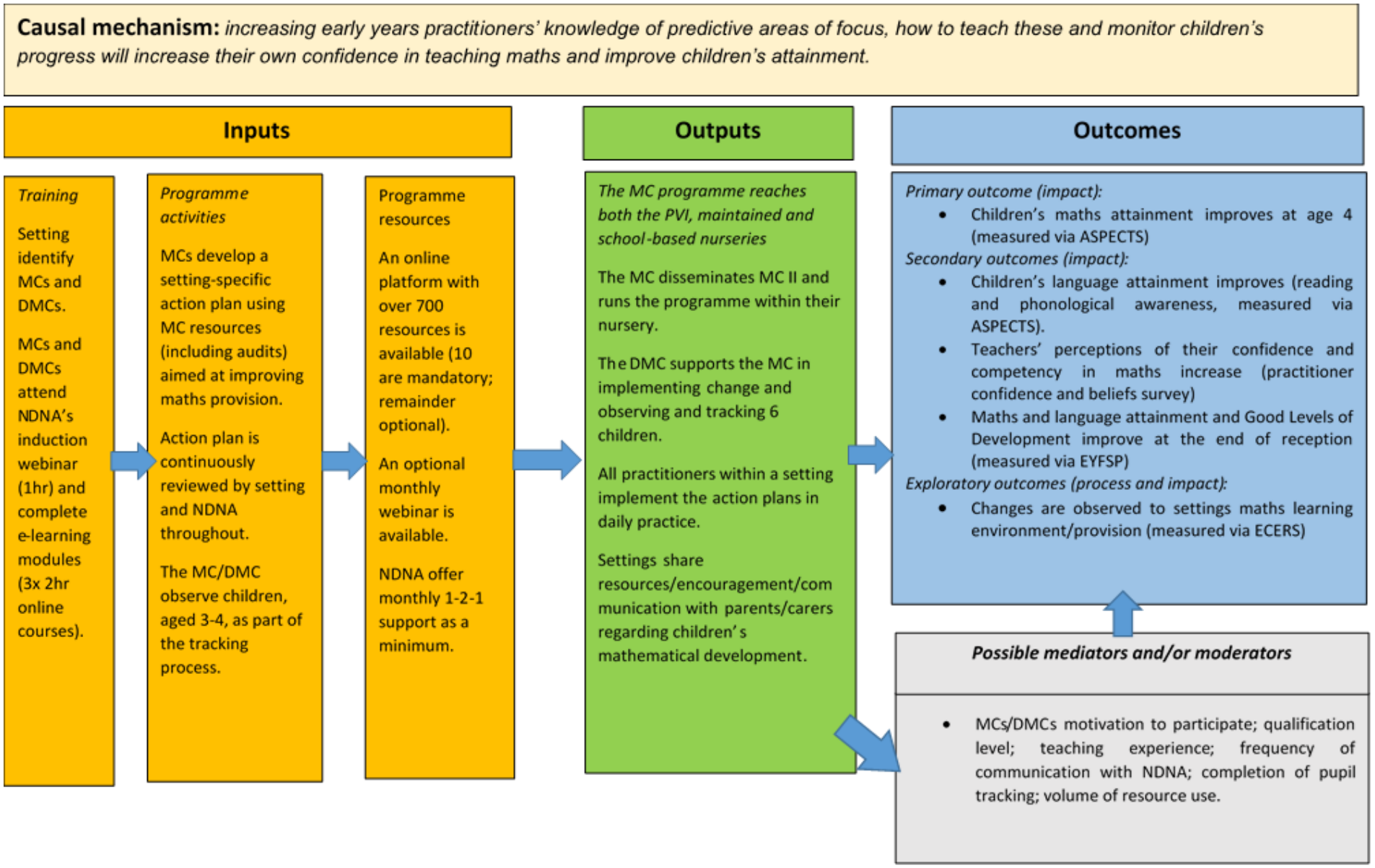
A detailed theory of change was originally developed by Evangelou and Mathers (2018) as part of the first MC effectiveness trial. The logic model below (Figure 1) includes the core components in respect of inputs, outputs, outcomes, and potential mediators and moderators. It was developed by the evaluation team on advice from the delivery team and builds on the work of Evangelou and Mathers (2018) while also considering key changes to the programme since the first MC effectiveness trial (detailed in Appendix C).

The causal mechanism of the logic model is that *increasing early years practitioners' knowledge of predictive areas of focus, and how to teach these and monitor children's progress, will increase their own confidence in teaching maths and improve children's attainment.*

The Early Intervention Foundation recognises the importance of creating high quality provision via high-quality training (Asmussen et al., 2018). It also identifies that between the ages of three and five is the ideal time to rectify income-related learning gaps in children's understanding of numbers. Preschools that helped children to understand early number concepts reported better outcomes in maths and overall later achievement (Mullis et al., 2012). The inputs of the MC programme aim to build confidence and professional understanding of teaching early years maths among practitioners, which are recognised issues within early years teaching (All Party Parliamentary Group for Maths and Numeracy, 2014). These inputs include the MC and DMC participating in relevant training (for example, online induction, modules, and webinars) that aim to equip the practitioners with a comprehensive understanding of the main areas of early years mathematical learning. Alongside this, practitioners implement programme tools including nursery-level action plans, downloadable trackers to monitor child progress, and online resources. These underpin the programme's outputs and enable the MC or DMC to evaluate existing practice, disseminate new learning to other practitioners within

their nursery, and change current practice. Together, these outputs aim to embed and increase the frequency and quality of maths routines, activities, exchanges, and interactions in daily early years practice (Frye et al., 2013). At the child level, this was intended to lead to improvements in children's maths attainment with spillover effects into language. Such spillover effects would be expected due to the implied causal relationship between language and mathematical development (Chow et al. 2021): children mainly learn mathematics via verbal instruction. Therefore, as practitioners' language and explanations of new mathematical concepts improves in complexity and frequency as a result of the MC programme, so too would children's language attainment. At the staff level, this would increase staff confidence in teaching early years maths and improve maths provision and the learning environment. It was anticipated that engagement with the programme may depend on practitioner's qualifications and experience.

Figure 1: Logic model



Evaluation objectives

The latest version of the protocol and statistical analyses plan (SAP) for the current effectiveness trial can be found on the EEF project website.

Effectiveness trial impact evaluation

Research questions

RQ1 What is the impact of the MC programme, in comparison to usual early years nursery provision, on the maths skills of preschool children aged three to four (primary outcome)?

RQ2 How effective is the MC programme at improving nursery practitioners' confidence in supporting children's maths development in comparison to usual early years nursery provision (secondary outcome 1)?

RQ3 What is the impact of the MC programme, in comparison to usual early years nursery provision, on the development of language (reading and phonological awareness) of preschool children aged three to four (secondary outcome 2)?

RQ4 What is the feasibility of accessing Ages and Stages Questionnaire (ASQ-3) data completed when children were two years old from NHS Digital? And how does this data correlate to maths and language development at three and four years old (measured using Assessment Profile on Entry for Children and Toddlers, 'ASPECTS')?

Effectiveness trial implementation and process evaluation

Research questions

RQ1 Is the MC programme delivered to MCs and DMCs with fidelity within both PVI nurseries and SNs?

1.1. Are nominated staff (MCs, DMCs) accessing the available e-learning modules and the support as specified in the programme plan?
1.2. How effective and appropriate is the level of support and training (for example, content, coverage, dosage, and duration) for MCs and DMCs?
1.3. What are the barriers for MCs and DMCs to engage with the e-learning modules?
1.4. What are the necessary conditions (facilitators) for MCs and DMCs to engage with the e-learning modules and the one to one support?

RQ2 To what extent is the MC programme implemented as planned within nurseries?

2.1. Do MCs and DMCs adhere to their roles as specified in the programme?
2.2. Do nursery practitioners implement the agreed action plans in their daily practice?
2.3. What are the barriers for MCs, DMCs, and practitioners to implement MC in their classroom practice?
2.4. What are the necessary conditions for nursery practitioners to implement MC into practice?

RQ3 What are the different stakeholders' viewpoints on the MC programme?

3.1. What are the perceived impacts of the MC programme on nursery practitioners' classroom practice in general?
3.2. What are the perceived impacts of the MC programme on nursery practitioners' math-related classroom practice, practitioners' confidence in teaching children maths, and practitioners' beliefs about children and maths?
3.3. What are the observed impacts on children's maths attainment?
3.4. How can the MC programme be improved?
3.5. What are the observed impacts of the MC programme on nursery practitioners' maths-related classroom practice?

RQ4 To what extent does the MC programme impact evaluation process adhere to the plan?

4.1. Do nursery MCs and DMCs meet the specified recruitment criteria for the MC programme?

4.2. Does children and family recruitment process adhere to the recruitment strategy?
4.3. Do baseline and outcome test administrators (teachers or independent research assistants) effectively and appropriately evaluate children's maths attainment?
4.4. Are there any sample attrition effects, and how might they affect the estimates of the impact of the MC programme?

RQ5 What is 'usual practice' in all sample nurseries?

RQ6 What maths-related professional development (PD) opportunities do staff have in control group nurseries?

6.1. What are the perceived impacts of these maths-related PD opportunities on nursery staff's maths-related classroom practice?
6.2. What are the perceived impacts of these maths-related PD opportunities on children's maths attainment?
6.3. What other maths-related PD opportunities are nursery staff looking for?

RQ7 What is the perceived impact of the COVID-19 pandemic on the delivery of the MC programme?

Ethics and trial registration

The University of York, Health Sciences Research Governance Committee (HSRGC) granted ethical approval for the trial on 29 November 2019 and approved a substantial amendment to the trial on 10 November 2020, which primarily covered changes to the trial as a result of COVID-19 (changes are summarised in the Methods section of this report). The School of Education Ethics Committee at Durham University were informed of the trial and the substantial amendment.

A memorandum of understanding (MoU) was signed by all nurseries to cover the requirements of the project (see Technical Notes). Participating nurseries also signed a data sharing agreement (DSA) issued by the University of York, and an agreement with the Centre for Evaluation and Monitoring (CEM) at Cambridge Assessment to cover the use of the ASPECTS software (End User Licence Agreement; EULA).

The trial was registered on 12 February 2020: ISRCTN31930534.

Data protection

All data collected for the trial was treated with the strictest confidence and processed and stored in compliance with the General Data Protection Regulation (GDPR) and Data Protection Act, 2018. The University of York was the data controller and a data processor, as defined in the GDPR. Other data processors, whom relevant data were shared with for the purposes of the trial included Durham University (to carry out the IPE), A+ Education (to complete ECERS assessments), and CEM at Cambridge Assessment (for administering the ASPECTS assessment). Once the data has been submitted to the Office for National Statistics Secure Research Service (ONS SRS) for archiving in the EEF data archive, the EEF will hold data controller responsibility.

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

Nurseries were provided with information regarding data protection and data sharing for the trial within the three documents mentioned above: the MoU (see Technical Notes), the DSA (issued by the University of York), and the EULA (issued by CEM at Cambridge Assessment for use of the ASPECTS assessment). Likewise, parents/carers of eligible children were informed about data protection and data sharing for the trial via an information sheet (see Technical Notes) distributed by nurseries. Parents/carers had to complete and return a consent form to allow data about their child to be shared for the purposes of the trial and to allow their child to complete assessments for the trial.

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

was also stored on Durham University's One Drive for Business cloud service with access restricted to members of the evaluation team.

Electronic transfers of personal child data between nurseries and the York Trials Unit, University of York, evaluation team were completed via encrypted spreadsheet sent through the University of York DropOff facility (a secure file transfer service). Paper consent forms were collected from nurseries and returned via courier to the University of York, where they were stored in locked filing cabinets in a room with restricted access. eConsent forms and survey data were collected using Qualtrics survey software. ASPECTS assessment data was collected and stored online via CEM's (Cambridge Assessment) servers. York Trials Unit had access to each nursery's ASPECTS account and CEM provided assessment data to York Trials Unit via secure file transfer.

The dataset for statistical analysis was anonymised. No nurseries, staff members, or children are identifiable in the report nor will be identifiable in the dissemination of any results. Electronic data and paper documents including identifiable personal child data will be securely archived and disposed of by York Trials Unit five years after the end of the study (2028). Identifiable personal data about adult data subjects (for example, nursery staff) will be kept for five years after the end of the study (2028). Anonymised electronic data and paper documents will be kept indefinitely by the evaluation team and may potentially be shared with other research teams or archiving organisations.

Project team

Evaluation team

University of York, York Trials Unit

Dr Lyn Robinson-Smith, Assistant Professor in health and education research. Lyn is experienced in leading and delivering large trials, particularly in the early years. She is the joint principal investigator and was responsible for the design of the impact evaluation and oversight and management of the trial. Lyn was also responsible for providing training for the baseline and post-test outcome assessments, led the recruitment of nurseries to the IPE, and led the writing the final report.

Hannah Ainsworth, an experienced education and health care trial manager. She was the joint principal investigator and was responsible for the design of the impact evaluation and oversight and management of the trial until August 2022.

Professor David Torgerson, co-director of York Trials Unit. David Torgerson has worked on numerous RCTs, including many in education and the social sciences. He supported the design and conduct of the trial. He took over as joint principal investigator in August 2022.

Professor Catherine Hewitt, senior trial statistician and co-director of York Trials Unit with experience working on numerous RCTs including educational trials. She provided input into the statistical analysis.

Louise Elliott, Louise Elliott has worked on a large number of EEF trials and has been involved in trial coordination, data management, and coordinating testing. She was responsible for the data management aspect and assisted in delivering testing on the trial.

Caroline Fairhurst, senior statistician, currently supporting a number of trials including several EEF-funded trials, within York Trials Unit. She oversaw and undertook the statistical analysis, contributed to the writing of the final report, and will take responsibility for archiving data with the FFT.

Kalpita Baird, a statistician, currently supporting a number of trials within York Trials Unit. She assisted with the development of the statistical analysis plan and child and nursery level randomisation.

Dr Katie Whiteside, a trial coordinator who has worked on a number of RCTs evaluating education and healthcare interventions. Katie coordinated the child recruitment, baseline, and post-test outcome assessment phases, and contributed to writing the final report.

Danielle Podmore, a trainee statistician currently supporting a number of trials within York Trials Unit. She contributed to the writing of the statistical analysis plan, undertook the statistical analysis, and contributed to the writing of the final report.

Heather Leggett, an applied researcher in the York Trials Unit with experience of conducting quantitative and qualitative research across a range of projects in public health and education. Heather contributed towards the qualitative analysis of the IPE survey data and report writing.

Tom Davill, the Trial Support Officer for the evaluation. Tom contributed to supporting nurseries through the child recruitment and baseline testing stages and assisted in delivering outcome post-testing.

University of York, Department for Education

Professor Carole Torgerson, an expert in RCT design and conduct, has been the principal investigator or a co-investigator on over 25 RCTs. Carole was principal investigator at Durham University until October 2020 before moving to the University of York. She contributed to the overall design and conduct of the impact evaluation, led the design and analysis of the IPE, and contributed to writing the final report.

Durham University

Dr Xiaofei Qi, Assistant Professor at Durham University and an associate of the Cambridge Psychometrics Centre. Her substantive area is early years. She contributed towards the content of the IPE surveys and interviews and conducted elements of the IPE data collection and analysis. She contributed to the report writing.

Vic Menzies, an experienced education trial coordinator and researcher with a particular focus on maths development and learning. She contributed expertise to the design and conduct of the evaluation, particularly the IPE. Vic was Principal Investigator at Durham University from October 2020.

Delivery team

Stella Ziolkowski, National Day Nurseries Association. Stella Ziolkowski is Director of Quality and Training at the NDNA. She had overarching contract responsibility for the delivery of outcomes and milestones, reporting to the EEF, and final approval of deliverables, processes, and procedures in relation to the trial.

Dr Sue Gifford, Roehampton University. Sue Gifford is a specialist maths adviser and provided mathematical advice for programme content.

Paula Dunn, NDNA. Paula Dunn is the MCs lead and was responsible for providing nurseries with their induction to the programme and continued one to one support for Champions.

Freya Roper, NDNA. Freya Roper was project manager with contract management responsibility for the day to day delivery of the trial, including milestones tracking, the recruitment process, and financial monitoring.

Kathryn Moses, NDNA. Kathryn Moses is a project officer and assisted with programme coordinating and recruitment, record keeping, and tracking mandatory outcomes for delivery.

Fiona Bland, NDNA. Fiona Bland is an early years advisor who was the MCs support adviser within the trial and covered for staff absences.

Methods

Impact evaluation design

This effectiveness trial was a two-armed, cluster randomised controlled trial with random allocation at the nursery level. Table 3 summarises the trial design.

Nursery-level randomisation was used to avoid potential contamination between groups. Nurseries were randomly allocated 1:1 to one of two groups:

- intervention—nurseries allocated to receive the NDNA MC programme plus usual nursery provision; or
- control—nurseries allocated to continue with usual nursery provision.

All nurseries (that is, in both intervention and control) received £250 after parent/carer recruitment, before pre-testing, and £250 after completing the outcome post-testing.

Table 3: Evaluation design

Trial design, including number of arms		Two-armed cluster randomised controlled trial
Unit of randomisation		Nursery
Minimisation factors		Nursery type (2 levels: PVI; SN and maintained nurseries). Nursery size (2 levels: < 30, which was the median number of children leaving for primary school in 2022 at participating nurseries; ≥ 30). Number of staff at the nursery holding a degree qualification in early years (2 levels: 0 graduates; ≥ 1 graduate).
Primary outcome	**Variable**	Child maths attainment after 7 months intervention exposure.
	Measure (instrument, scale, source)	ASPECTS maths attainment score, 0–29, Centre for Evaluation and Monitoring (CEM) at Cambridge Assessment.
Secondary outcome(s)	**Variable(s)**	Practitioner confidence (in teaching children maths) after 7 months' intervention exposure. Child Language attainment after 7 months' intervention exposure. Child development at 2 years old and its correlation to child development at 3 and 4 years old. Longitudinal: child attainment at the end of reception year at school (data to be obtained in November 2023 and published in an addendum report in spring 2024).
	Measure(s) (instrument, scale, source)	Practitioner confidence: maths. Adapted 'Early Math Beliefs and Confidence Survey' by Chen et al. (2014). Only the adapted subscale 'confidence in helping nursery aged children learn maths'. ASPECTS language (reading and phonological awareness) score, 0–53, CEM at Cambridge Assessment. ASQ-3 (Squires and Bricker, 2009) at 2 years old, data gathered via NHS Digital and its correlation to ASPECTS (described above).

		Longitudinal: Early Years Foundation Stage Profile data (completed at the end of reception) collected from National Pupil Database in November 2023 (to be published in an addendum report in spring 2024).
Baseline for primary outcome	**Variable**	Child maths attainment.
	Measure (instrument, scale, source)	ASPECTS maths attainment score, 0–29, CEM at Cambridge Assessment.
Baseline for secondary outcome	**Variable**	Child language attainment.
	Measure (instrument, scale, source)	ASPECTS language (reading and phonological awareness) score, 0–53, CEM at Cambridge Assessment.

Due to COVID-19, the timeline of the effectiveness trial was amended within the protocol (as outlined in the protocol version history table; Robinson-Smith et al., 2022). Initially, the effectiveness trial was due to run within nurseries over the 2020/2021 academic year, but this was delayed by one year until 2021/2022. As outlined further in Appendix E, the pilot trial evaluation, planned to run between January and June 2020, had to be paused in March 2020 and was heavily impacted by the COVID-19 lockdowns.

Mindful of the impact of COVID-19 within nurseries, a number of other changes were made to the effectiveness trial methodology. As the timeframe for recruitment and baseline assessment was initially expected to be shorter for SNs (as, compared to PVI nurseries, all children would likely not have been present in the SN until the beginning of the academic year), it was initially planned that an independent research assistant (RA) would visit all SNs to complete the baseline ASPECTS assessments unless the SN had requested to complete these itself. For PVI nurseries, it was always intended that a practitioner from each nursery would be asked to complete the baseline ASPECTS assessments; the protocol was revised to take this approach with all nurseries to minimise the need to visit nurseries (when this decision was made, nurseries were typically restricting external visitors due to COVID-19). There was, however, provision to send an assessor to complete baseline assessments in participating nurseries that were unable to complete assessments within the agreed timeframe and could accommodate an assessment visit. The protocol was also updated to allow for parent/carer consent for participation in the evaluation to be collected online (eConsent) as an alternative to using paper consent. eConsent presented a number of advantages, including allowing all nurseries (not just PVI nurseries) to start the parent/carer consent process prior to the end of 2020/2021 and run it over the summer holidays. Using eConsent also meant that parents/carers of all eligible children could be invited at the same time regardless of whether they were in the nursery over the summer holidays or were due to start in September 2021.

Another change to the protocol due to COVID-19 was the addition of an IPE research question (RQ7) in order to evaluate the perceived impact of the pandemic on the delivery of the programme within the effectiveness trial. An additional research question (RQ4) for the impact evaluation was also added in relation to exploring the feasibility of accessing ASQ-3 (Squires and Bricker, 2009) data, completed when children were two years old, from NHS Digital and how this correlated to maths and language development at three and four years old (measured using ASPECTS). The collection of ASQ-3 data and its correlation with ASPECTS was intended to be explored within the pilot study but this was impacted due to COVID-19 (see Appendix E for further discussion) and so this exploration was planned to be incorporated within the effectiveness trial instead. Unfortunately, collection of ASQ-3 data was ultimately determined to be infeasible, as detailed in the Results chapter, and so analyses of ASQ-3 data were not possible.

Participant selection

Participants

Nurseries

The delivery team led the recruitment of nurseries to take part in the trial, supported by the evaluation team. Recruitment began in January 2020 but was paused between March and December 2020 due to the COVID-19 pandemic and then recommenced in January 2021. Recruitment strategies included a page dedicated to the trial on the NDNA's website, emails to nurseries in recruitment areas, marketing through social media channels, promotion via sector press and public relations work, working with contacts in targeted local authorities and providing them with recruitment materials to push out at a local level, and working with Early Education to promote the trial.

Nursery eligibility criteria:

- PVI providers based on non-domestic premises, maintained nursery schools or children's centres, or government funded infant or primary SNs providing nursery provision for three- and four-year-olds (who were due to begin reception in September 2022);
- nurseries with a minimum of 15 children aged three to four who attended for a minimum of 15 hours a week and were due to begin reception in September 2022;
- nurseries not currently using the NDNA MC programme and had not done so in the past;
- nurseries not currently taking part in the evaluation of the Department for Education's Early Years Professional Development Programme or any other early years trial funded by the EEF or similar funder;
- nurseries that agreed to all requirements outlined in the Information for Nurseries and MoU document, DSA, and CEM's EULA;
- only one participating nursery per nursery chain or academy trust;
- nurseries that accept government childcare subsidies (that is, 30 hours free childcare for three- and four-year-olds); and
- nurseries located in England.

Recruitment for this trial was initially focused on nurseries within the East Midlands and West Midlands. However, in order to meet recruitment targets, nurseries from other regions across England were later recruited (the main recruitment areas included the West Midlands, Oxfordshire, Northamptonshire, Greater London, West Yorkshire, Greater Manchester, Cheshire, Tyne and Wear, Somerset, and Hampshire). The research team planned to recruit approximately 138 nurseries: 69 in each of the intervention and control arms—approximately 96 PVI and 42 SNs, representing 70% PVI and 30% SN in line with national provision as, excluding childminders, SNs make up 27% of early years providers in England (DfE, 2019a).

Interested nurseries were requested to complete an expression of interest survey accessed via a link on the NDNA's website. The NDNA also shared the expression of interest survey via the NDNA delivery information sheet, e-bulletins, and specific emails to nurseries. This covered the nursery eligibility criteria and collected nursery regional location and contact details. Eligible nurseries were then sent a study information sheet and MoU, which provided full details relating to a nursery's involvement within the trial. Nurseries that returned a MoU prior to the recruitment pause due to COVID-19 were sent the updated MoU after recruitment recommenced. Nurseries that were willing to participate returned a completed and signed MoU. Nurseries were also required to sign a DSA issued by the University of York and EULA for use of CEM's ASPECTS assessments before being recruited into the trial.

The delivery team held information sessions over Zoom for nurseries that had been sent a MoU and a representative from the evaluation team also joined these sessions. The purpose of the information sessions was to ensure settings made an informed choice about joining the study; this was integral to help settings understand the commitment they were making.

In addition, the delivery team followed up with nurseries that had been sent a MoU via phone call to check understanding of the trial and MoU, answer questions, check eligibility, and encourage participation.

Nurseries were recruited on a first come first served basis, with location taken into consideration when following up with nurseries (as this was advantageous when scheduling post-testing). Nurseries that returned trial paperwork after 138 nurseries had been recruited were placed on a reserve list to be called upon should a recruited nursery drop-out before randomisation (reserves were included in the study, where necessary, subject to there being sufficient time for the new nursery to get through recruitment and baseline data collection before the randomisation period closed).

All nurseries—both intervention and control—received £250 (bank transfer from the delivery team) after parent/carer recruitment, before baseline pre-testing. This thank you payment was to allow nursery staff time to be freed up for baseline testing. All nurseries received £250 (bank transfer from the delivery team) after completing the outcome post-testing.

Children

Child eligibility criteria:

- children aged three to four (or who were due to turn three before September 2021) starting reception in September 2022;
- children who attended or were due to attend nursery for a minimum of 15 hours per week during term time from September 2021;
- children whose parents/carers anticipated that they would remain at the nursery for the duration of the trial—that is, they did not foresee they would leave the nursery until June 2022; and
- children who completed the ASPECTS baseline assessment.

Children were not eligible to take part in the trial if practitioners considered them to have significant Special Educational Needs or Disabilities or English as an Additional Language, where an extreme language barrier existed which would have prevented them from accessing the ASPECTS assessment or they would have been distressed through completing the assessment.

In July 2021, the evaluation team provided each nursery with a PDF copy of the Parent/Carer Information Sheet (see Technical Notes) and a link for parents/carers to securely complete a consent form online (eConsent) using Qualtrics survey software. Nurseries were requested to email these to the parents/carers of all children who met the first three eligibility criteria. Nurseries were encouraged to begin the parent/carer consent process over the summer prior to the start of the 2021/2022 academic year, contacting parents/carers of children who were on the school's pre-registration lists via email. Nurseries were asked to inform the evaluation team if they preferred to send out and collect paper consent forms or use a combination of eConsent and paper consent (for example, if they thought that some parents/carers would prefer to receive paper copies). In that case, the evaluation team either emailed the nursery a copy of the parent/carer consent form and information sheet to print and distribute or posted copies to the nursery, depending on the nursery's preference. The evaluation team arranged for paper consent forms returned to the nursery to be collected via courier and returned to the University of York.

Child recruitment closed in each nursery once it had obtained parent/carer consent (paper consent or eConsent) for at least ten eligible children to participate in the evaluation and the parents/carers of all eligible children had been given sufficient opportunity to compete a consent form. Nurseries that recruited between six and nine children and had exhausted recruitment progressed to participation in the trial in order for the number of recruited nurseries/children to be as close as possible to the recruitment target. On the eConsent/paper consent form, parents/carers were requested to provide consent for their:

- child to participate in the evaluation;
- child to complete the baseline and outcome post-test assessments;
- child's nursery to provide the evaluation team with data regarding their child as outlined in the information sheet (see Technical Notes) and be contacted for this information should their child's nursery be unable to provide it; and
- consent for long term tracking of their child's educational outcomes through the National Pupil Database for the purposes of the evaluation.

On the consent form, parents and carers were also asked if their child was eligible to receive the 15 hours of government funded childcare at age two and, if eligible, were asked to indicate whether or not they utilized these funded nursery hours (a child may be eligible for the childcare funding at two years old, but their parent/carer may choose not to send them to nursery).

In cases where nurseries gained parent or carer consent for ten or fewer children, all children were selected for baseline and post-testing. In cases where nurseries gained consent for more than ten children, a sample of ten were randomly chosen for baseline and subsequent post-testing. Where possible, the evaluation team aimed to include at least one EYPP child per nursery to have adequate power to conduct analyses in the EYPP subgroup (see the Sample size section below). Therefore, we randomly selected up to three eligible children with EYPP status (or all of them if there were three or fewer eligible and including more than one where this was possible, which allowed for some attrition at follow-up), then we randomly sampled from the remaining, unselected children (EYPP and non-EYPP) to make up the ten. The remaining unselected children (EYPP and non-EYPP) were randomly ordered and were considered as reserves to be asked to complete baseline testing if one of the ten selected children did not want to complete the baseline assessment or were absent over the baseline testing period. Baseline data collection took place before the nursery was randomised and informed of its allocation. Ultimately, some nurseries only had a small number of consenting children, and the decision was made to only randomise nurseries if six or more consent forms were returned.

Regardless of whether or not a consented child completed the assessment, all children received the intervention if their nursery was randomly allocated to receive it.

Outcome measures

Primary outcome

ASPECTS (Assessment Profile on Entry for Children and Toddlers), developed by CEM and hosted by Cambridge Assessment, was the primary baseline and post-test outcome measure. ASPECTS has been specially designed for children aged three to five years old and is aligned with the elements of the EYFS 'prime and specific areas of learning and development'. Early maths skills that are assessed include digit identification, counting, shapes, number problems, and ideas about maths—areas targeted by the MC programme, which provides a holistic approach to improving maths attainment. The early maths skills component of ASPECTS formed the primary outcome measure within the first MC trial (Robinson-Smith et al., 2018), which allows for a comparison of results. Participating children were assessed using ASPECTS at baseline at the start of 2021/2022 (before their nursery was randomised) and again at the end of that academic year for outcome post-testing. The evaluation team aimed to assess ten children per nursery for whom parent/carer consent was received, at baseline and again at the post-test outcome timepoint. The evaluation team liaised with nurseries to arrange the baseline and post-test outcome assessments.

ASPECTS is a child-friendly, computer-based assessment designed to be used on a one to one basis with children. The programme asks children to complete a series of activities and an adult submits the child's responses on the computer. At first, the child is asked to write their name, and this is scored by the adult against examples. The software then plays an audio recording of a story to the child and asks a number of questions. While all children hear the same story, ASPECTS adopts an adaptive design whereby the questions asked are dependent on the child's responses (for example, more challenging questions are provided when a child answers a question correctly). ASPECTS uses Rasch measurement to estimate the item difficulties. All items are categorised, and more difficult items from each category are no longer presented once the child has made a certain number of mistakes in that category. The early maths subscales of the measure, which take approximately 10 to 12 minutes per child, were used at baseline and post-testing. The maths score (range 0 to 29) was the primary baseline and post-test outcome measure; a higher score indicates greater attainment.

At baseline, prior to randomisation of the nursery, a teacher or practitioner from each nursery (a child's 'key worker') was asked to complete ASPECTS with the participating children. As children were very young at baseline (typically three), having a familiar nursery practitioner or teacher administer the assessment was thought to help the children to perform to the best of their ability and minimise missing data. We requested that the same staff member should complete ASPECTS with all participating children at the nursery, if possible. We provided nurseries with training in how to set up and administer ASPECTS via an online, 15-minute, pre-recorded webinar as well as accompanying written instructions. We added the names and details for the (up to) ten children selected for baseline assessment to each nursery's online ASPECTS account before providing each nursery with their login details. Nurseries were asked to contact us if one of

the selected children did not want to complete the baseline assessment or were absent over the baseline testing period. In these cases, and if a nursery had gained parent/carer consent for more than ten children, we added the details for the next available child from the list of 'reserve' children who had not been initially selected for baseline assessment. There was provision for the evaluation team to send an assessor to complete baseline assessments in participating nurseries that were unable to complete assessments within the agreed timeframe.

At the time of outcome post-testing, as far as possible, ASPECTS was administered in all nurseries by independent, blinded research assistants (RAs). All RAs were employed specifically for this role: they received training from the evaluation team, had an enhanced DBS check, and underwent relevant safeguarding and data protection training. We advised nurseries that a child's key worker or familiar staff member should be available to chaperone the assessment conducted by the RA to ensure the child felt comfortable. Each nursery received at least one assessment visit from a RA and, where feasible, nurseries received additional visits as necessary to try to assess all participating children. RAs were reminded not to ask which group the nursery they were visiting was randomly allocated to. If an RA became unblinded (for example, the nursery mentioned that they had been following the Maths Champions programme), they were asked to email the evaluation team so this could be recorded (but there were no known instances of this). Where a blinded RA was not available to visit a nursery (for example, due to illness or rails strikes), or in cases where it was not feasible to send another RA for a second or third visit to finish the assessments (for example, if the nursery was remote or if the remaining children to assess had poor attendance), we allowed for post-test outcome data to be collected by an unblinded, independent assessor (for example, a member of the evaluation team) or a practitioner or teacher within the nursery in order to minimise attrition.

In cases where children had moved to a new nursery before outcome post-testing, the evaluation team planned to follow up such children and assess them in their new nurseries (provided we had agreement from new nurseries). The proportion of children for whom this strategy was to be employed was dependent on the numbers of children identified as having moved to new nurseries. The aim was to achieve a low level of attrition overall weighed against the cost implications of assessing in new nurseries. It was planned that new nurseries that facilitated the outcome assessment would receive £100. During spring term 2022 and again at the start of summer term 2022, we contacted each participating nursery to check whether each participating child, who completed the baseline assessment, still attended their nursery and, if not, to provide the name of their new nursery (parents/carers had agreed to this on the consent form).

Secondary outcomes

The literacy and language score from ASPECTS, carried out at baseline and post-test, was a secondary outcome. Early literacy skills that are assessed include reading and phonological awareness. This is scored from 0 to 53, where a higher score indicates greater attainment. The MC programme aims to increase the frequency of use of maths terminology between practitioners and children in all interactions; therefore, there is potential for intervention spillover effects in the domain of literacy and language. The literacy and language component of ASPECTS formed a secondary outcome measure within the first MC trial (Robinson-Smith et al., 2018), which allows for a comparison of results.

Practitioner confidence (in teaching children maths), assessed using a short online survey adapted from Chen et al. (2014), was also a secondary outcome. Increasing practitioners' confidence in using maths is a key focus of the MC programme. We requested for the survey to be completed by the nominated MC or DMC in intervention nurseries and comparable staff in control nurseries. The survey was completed at post-intervention only. The original survey consists of three subscales: 'beliefs about nursery aged children and maths' (eight items), 'confidence in helping nursery aged children learn maths' (11 items), and 'confidence in own maths abilities' (nine items).

The three subscales from the original survey by Chen et al. (2014) each produce separate scores and cannot be combined. All three subscales were collected and analysed in the first MC effectiveness trial but after careful consideration only subscale two was used for this trial. As per the reasoning behind this decision, in the first MC trial, the intervention was designed to improve practitioners' own maths abilities (which could justify the inclusion of subscale three) while in this second trial, improving practitioners' own maths abilities is no longer a focus of the intervention. Moreover, the evaluation and delivery teams agreed that there were limitations in using subscale one since there could be debate about what constitutes a 'correct' or 'better' belief, which would make interpretation of a difference in scores between the randomised groups challenging. Also, some questions asked about beliefs about the characteristics of incoming children, which the intervention would not be expected to change. Therefore, only the second subscale, 'confidence in helping nursery aged children learn maths', was used. Practitioners were asked to rate their agreement

with each item on a Likert scale, from strongly disagree to strongly agree. Each item was scored from one to five. Scores for items in the subscale was summed to produce a summary score ('confidence in helping nursery aged children learn maths' scored from 11 to 55). Practitioner confidence using the Chen et al. (2014) survey, including subscale two, formed a secondary outcome measure within the first MC trial (Robinson-Smith et al., 2018), which allows for a comparison of results.

The ASQ-3 (Squires and Bricker, 2009) is used to capture the skills and development of children at two years old. The domains of the ASQ-3 include communication, gross motor, fine motor, problem-solving, and adaptive skills. A score is assigned to each development domain. Within any screened domain, less than two standard deviations below the mean area score is considered a positive screen. The ASQ-3 is validated and standardised and has been reported to be accurate in detecting developmental delays in children. The ASQ-3 is used routinely by health visitors who request parents complete a 15-minute questionnaire as part of a health check when their child is two years old. The data from the questionnaire is stored, and accessed, via NHS Digital (Public Health England, 2018; 2020).

Sample size

From protocol

The following assumptions were made:

- a nursery-level intra-cluster correlation coefficient (ICC) of 0.17, derived from the first MC trial (Robinson-Smith et al., 2018);
- a pre-test/post-test correlation of 0.59 (from the first MC trial);
- ten children per nursery at baseline; and
- a one to one allocation at the nursery level.

Based on 138 nurseries (1,380 children), we would have 80% power to show an effect size of 0.20 of a standard deviation between the control and the intervention groups, allowing for 15% attrition at the child level (Table 4).

This is calculated as follows. Assuming 1,380 children are randomised, and there is 15% attrition at post-test, there will be 1,173 children included in the analysis. These 1,173 children are spread across 138 nurseries, which in a cluster trial equates to an approximate effective sample size in an individually randomised trial of 516. This is obtained by dividing the sample size by the design effect of $1 + ((m - 1) \times \rho)$ where m is the average cluster size at analysis and ρ is the ICC. Altogether, this equates to $1{,}173/[1 + ((8.5 - 1) \times 0.17)]$ (Rutterford, Copas and Eldridge, 2015). We assume the correlation between the ASPECTS maths score measured at pre- and post-test is 0.59; therefore, an analysis adjusting for baseline score, as we plan here, will have the same power with 516 children as a t-test comparing two equal-size groups with a total of $516/(1 - 0.59^2)$ children (Borm, Fransen and Lemmens, 2007). Stata v17 was used to estimate the MDES based on a t-test comparing two groups with a total of 792 children using the command `power twomeans 1, sd(1) power(0.8) n(792)`, which gives 0.20.

A subgroup analysis was proposed for the primary outcome among children eligible for EYPP. Owing to the proposed sampling strategy of eligible children to participate in the trial, we anticipated to have at least one EYPP child from each nursery included in this analysis (though it was possible that some nurseries had no children with EYPP). If most nurseries only had one child eligible for EYPP who contributes to this analysis then we planned for the analysis to be conducted at the nursery level, aggregating child outcomes by taking the mean for eligible EYPP children in that nursery. Assuming a pre-test/post-test correlation of 0.59 (no design effect assumed since at nursery level), with 138 nurseries we would have 80% power to show an effect size of 0.38 of a standard deviation between the control and the intervention groups in the EYPP subgroup (Table 4).

If, however, more than half the nurseries had two or more eligible EYPP children who contribute to the analysis and the average number per nursery was great than or equal to two, then we planned to conduct this analysis at the child level accounting for the clustering by nursery. Assuming an ICC of 0.17, an average of two children per nursery at analysis, a pre-test/post-test correlation of 0.59, and 1:1 allocation at nursery level, we would have 80% power to show an effect size of approximately 0.30 of a standard deviation between the control and intervention groups in the EYPP subgroup (Table 6).

At randomisation

In total, 1,304 children were assessed using ASPECTS at baseline across 134 nurseries (average cluster size of 9.7). Assuming an ICC of 0.17, a pre-test/post-test correlation of 0.59, and 15% child-level attrition, we would have 80% power to detect an effect size of 0.20 between the two arms (Table 4).

For the EYPP analysis, at nursery level, the MDES would be 0.39 and at child level (assuming two children per nursery) the MDES would be 0.30 (Table 4).

Table 4: Sample size calculations

		Protocol		Randomisation	
		Overall	**EYPP***	**Overall**	**EYPP***
Minimum Detectable Effect Size (MDES)**		0.20	0.38 / 0.30	0.20	0.30 / 0.39
Pre-test/post-test correlations	**Level 1 (child)**	0.59	0.59	0.59	0.59
	Level 2 (nursery)	N/A	N/A	N/A	N/A
Intracluster correlations (ICCs)	**Level 2 (nursery)**	0.17	N/A / 0.17	0.17	N/A / 0.17
Alpha		0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two
Average cluster size		10	1 / 2	9.7	1 / 2
Number of nurseries	**Intervention**	69	69	66	66
	Control	69	69	68	68
	Total	138	138	134	134
Number of children	**Intervention**	690	69 / 138	638	66 / 132
	Control	690	69 / 138	666	68 / 136
	Total	1,380	138 / 276	1,304	134 / 268

* EYPP = Early Years Pupil Premium; figures either side of the / represent the two scenarios: (i) aggregating data to nursery level and (ii) conducting analysis at child level
** All estimates assume 15% child-level attrition at post-test.

Randomisation

Nurseries were randomly allocated, one to one, to either receive the MC programme (intervention group) or to continue with usual nursery provision (control group). A statistician at York Trials Unit, who had no involvement in the recruitment of nurseries and was blinded to their identity (nurseries were identified by a unique trial identifier), randomised nurseries using minimisation to ensure balance across the trial arms on nursery type, nursery size, and the number of graduate staff (see Table 3 for the levels of each minimisation factor). The median number of children leaving for primary school in 2022 was 30 and was calculated based on expected numbers from 138 nurseries that expressed interest in the trial. A dedicated computer programme, MinimPy (Saghaei and Saghaei, 2011), was used for randomisation.

In order to maximise the time that the intervention could be implemented in the nursery (if allocated to the intervention arm), rather than waiting for child recruitment and baseline assessments to be completed in all nurseries, nurseries were randomised as soon as possible after baseline tasks had been completed at their setting. It was logistically easier to wait until a number of nurseries were ready to be randomised before entering them into the MinimPy programme, so nurseries were randomised in 17 ‘batches’ between October and December 2021. Within each batch, the data (on the minimisation factors) for each nursery was entered into MinimPy in turn and the programme allocated them to a trial

arm. Naïve minimisation with base probability 1.0, following a random start, was conducted (that is, deterministic minimisation). This means that the allocation for the first nursery randomised was chosen by the programme at random (like tossing a coin) and after that the allocations for following nurseries are theoretically predictable based on ensuring the best possible balance between the groups on the specified minimisation factors. Naïve minimisation was deemed to be sufficient, however, given that the allocations were conducted in batches rather than prospectively, meaning predictability was not a concern and hence a random element was not required (Altman and Bland, 2005).

The final number of nurseries randomised into the trial was 134 (intervention 66; control 68). The trial statistician was not blinded to group allocation at analysis.

Statistical analysis

The statistical analysis followed the EEF's most recent guidance at the time (EEF, 2018). All analyses were conducted in Stata v17 (StataCorp, 4905 Lakeway Drive, College Station, Texas 77845 U.S.A.) on an intention-to-treat (ITT) basis, where data was available, using all nurseries and children in the groups to which they were randomised irrespective of whether or not they actually received the intervention.

Statistical significance was assessed using two-sided tests at the 5% significance level. Effect sizes based on the difference between the groups at the outcome post-testing is presented as an adjusted mean difference and as Hedges' g with 95% confidence intervals (CIs) and p-value and converted to an estimate for the number of months' progress.

The number of children identified as eligible for the evaluation, the number for whom parent/carer consent was received, the number selected to take part, and the number tested for ASPECTS at baseline and post-test outcome assessments is reported with reasons for non-participation where available.

The trial was designed and conducted, and has been reported, to CONSORT standards. A CONSORT diagram has been produced to show the flow of nurseries and children through the trial.

All outcome data has been summarised descriptively by trial arm.

Imbalance at baseline

Nursery- and child-level characteristics and baseline data are summarised descriptively by randomised group, both as randomised (to check the randomisation achieved balance) and as analysed in the primary analysis (to check whether attrition has introduced selection bias into the complete-case sample). Continuous measures are reported as a mean and standard deviation (SD) while categorical data is reported as a count and percentage. No formal statistical comparisons of baseline data were undertaken, except to report the differences in baseline scores (maths and language scores from ASPECTS) as a Hedges' g effect size and 95% CI.

Primary analysis

The early maths subscale of ASPECTS was assessed at baseline and post-test outcome timepoints. The maths score ranges from 0 to 29, with a higher score indicating greater attainment. The pairwise correlation between baseline and post-test outcome measurements for ASPECTS scores is presented. Histograms of pre- and post-test scores were produced.

Numeracy attainment for children in the intervention group and those in the control group was compared using a linear mixed model at the child level. Group allocation, baseline ASPECTS maths score, and nursery-level minimisation factors (nursery type—PVI or SN/maintained), nursery size, and number of staff at the nursery holding a degree qualification in early years) were included as fixed effects in the model. The continuous variables that were dichotomised to use as factors in the minimisation procedure (nursery size and number of graduate nursery staff) were included in their continuous form in the model.

Child-level fixed effects:

- baseline ASPECTS maths score (continuous).

Nursery-level fixed effects:

- number of staff at the nursery holding a degree qualification in early years (continuous);
- nursery size (continuous); and
- nursery type (PVI or SN/maintained; binary).

Adjustment was made for clustering at the nursery level by including nursery as a random effect and robust standard errors were specified to account for any potential heteroscedasticity.

Model equation

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 w_i + \beta_3 y_i + \beta_4 I_{Ti} + \beta_5 I_{Ai} + u_i + y_{ij}$$

Y_{ij} = response (post-test ASPECTS maths score) of the j-th of n_i members of the i-th cluster (nursery), *i=1,…,m, j=1,…,n_i*

m = number of clusters (nursery)

n_i = size of cluster (nursery) i

x_{ij} = baseline ASPECTS maths score for j-th member of i-th cluster (nursery)

w_i = number of staff holding a degree qualification in early years in i-th cluster (nursery)

y_i = size of i-th cluster (nursery)

I_{Ti} = indicator variable for type of i-th cluster (nursery) (0=PVI, 1= SN/maintained)

I_{Ai} = indicator variable for group allocation of i-th cluster (nursery) (0=Control, 1=Intervention)

β_0, β_1, β_2, β_3, β_4, β_5 = fixed effect parameters

$u_i \sim N(0, \varphi_b^2)$ = nursery-specific random effect and $\gamma_{ii} \sim N(0, \varphi_w^2)$ = individual-specific random effect

The normality of the standardised residuals was checked using a visual inspection of the QQ plot.

The primary analysis model included all post-test results regardless of the method of data collection (for example, via a blinded RA or unblinded practitioner or teacher at the nursery). This was explored further in a sensitivity analysis, described below.

The ICC associated with nursery for the primary outcome (both pre- and post-test) is presented alongside a 95% CI. The ICC at post-test was computed for the primary analysis model and also for an empty model—one without covariates. The ICC at baseline was calculated for a linear model with baseline as the outcome and nursery as a random effect.

Subgroup analyses

Subgroup analyses looking at gender, the average number of hours the child attends the nursery per week (dichotomised at the median number of hours), eligibility for EYPP, whether a child was eligible for FEEE at two years old, and PVI nurseries versus SN were undertaken for the primary outcome. These subgroup analyses were conducted by including the factor and an interaction term between the factor and allocation in the primary analysis model.

Whether a child was eligible for FEEE at two years old was assessed using the following three questions:

- Did the child receive funded childcare when they were two years old (nursery provided)?
- Was your child eligible to receive government funded childcare at two years old (from parent/carer consent form)?
- Did your child receive up to 15 hours government funded childcare at two years old (from parent/carer consent form)?

If any of these questions were answered ‘yes’ then we considered the child to be eligible for FEEE at two years old.

We also repeated the primary analysis for only the subset of participants eligible for EYPP. As stated in the sample size section, if more than 50% of the nurseries had only one child eligible for EYPP included in the model, then we planned to conduct this analysis at the nursery level whereby the pre- and post-test ASPECTS maths scores are averaged across children for any nursery with more than one child with EYPP status. Otherwise, if fewer than 50% of the nurseries had two or more eligible EYPP children who contribute to the analysis, then we planned to conduct the analysis at the child level as described for the primary analysis above. Ultimately, the latter approach was used (a similar approach to the primary analysis but within the restricted subsample of children eligible for EYPP).

Dosage

Intervention dosage is summarised and was defined as the length of time (in weeks) a nursery delivered the MC programme. In this effectiveness trial, the intended duration of programme delivery was seven to eight months. This started on the day the NDNA made contact with the nursery to begin the MC programme and ended when post-testing occurred or when the nursery expressed a desire to no longer implement the programme or when the NDNA withdrew its support, whichever occurred sooner.

Analysis in the presence of non-compliance

Compliance and fidelity were measured at the nursery level. Each nursery in the intervention arm was assessed for its implementation fidelity and compliance. This was measured by the NDNA which rated each nursery on compulsory (core) and optional aspects of the MC programme on a three-level scale: 2 = very engaged (‘green’), 1 = partially engaged (‘amber’), and 0 = not engaged (‘red’), as detailed in Table 5. This resulted in possible scores of 0 to 16 for core components with an additional 12 points for optional components.

For the purposes of this rating scale, in this particular trial with a multifaceted intervention, it was challenging to differentiate between compliance and fidelity, so this scale utilises information on both compliance and fidelity within one rating scale. The evaluation and delivery teams carefully considered, and agreed, the components that constituted engagement with the intervention, according to the logic model.

Table 5: Compulsory and optional components—compliance and fidelity rating

Criteria	Core or optional	Description	RAG rating
Identification of suitable MC (graduate or level 3 practitioner)	Core	MC with level 3 or graduate qualifications	Green = 2
		MC identified with < level 3 qualifications	Amber = 1
		MC with no level 3 qualifications or no MC identified	Red = 0
Identification of suitable deputy MC (qualified to at least level 3)	Core	DMC with level 3 qualifications or higher	Green = 2
		DMC with no level 3 qualifications	Amber = 1
		No DMC identified	Red = 0
MC and DMC complete induction	Core	MC and DMC complete induction	Green = 2
		Only MC or DMC complete induction	Amber = 1
		Neither MC or DMC complete induction	Red = 0
Completion by the MC of 2 courses: developing mathematical confidence in the early years: the big ideas of number sense; and developing mathematical thinking in the early years: shape space, measures and pattern—including characteristics of effective learning and sustained, shared thinking.	Core	Both completed	Green = 2
		One completed	Amber = 1
		Neither completed	Red = 0
Use of audit tool	Core	Audit tool used and audit completed	Green = 2
		Audit tool used but audit not completed	Amber = 1
		Audit tool not used	Red = 0
Completion and continued use of an action plan	Core	Action plan done and used as working document throughout	Green = 2
		Action plan done, started to be used but then not implemented	Amber = 1
		Action plan not done or not used	Red = 0

Use of up to ten mandatory resources provided through online platform: 3-4 year olds: build a maze, number hunt, delivering the post, mud kitchen, cars down a ramp, patterns, construction, tidy up time, snack time, outdoor games	Core	Use of at least 8 mandatory resources	Green = 2
		Use of 5–7 mandatory resources	Amber = 1
		Use of 4 or fewer mandatory resources	Red = 0
Engagement with one to one support provided by the NDNA	Core	Nursery always receptive to support from the NDNA	Green = 2
		Nursery sometimes receptive to support from the NDNA	Amber = 1
		Nursery never receptive to support from the NDNA	Red = 0
Possible total score core components			**16**
Track and monitor development of six children on termly basis.	Optional	All done and evidence uploaded	Green = 2
		Some done but needed support	Amber = 1
		None done	Red = 0
Monthly webinars	Optional	Attend two or more	Green = 2
		Attend one	Amber = 1
		Attend none	Red = 0
Completion by the DMC of two courses: developing mathematical confidence in the early years: the big ideas of number sense; and developing mathematical thinking in the early years: shape space, measures and pattern – including characteristics of effective learning and sustained, shared thinking.	Optional	Both completed	Green = 2
		One completed	Amber = 1
		Neither completed	Red = 0
Completion by MC/DMC of 'coaching as an educational lead' course	Optional	Both MC and DMC complete	Green = 2
		Only MC completes	Amber = 1
		Neither MC nor DMC complete, or DMC completes but MC does not	Red = 0
Reflection and completion of case study based on outcomes of action plan	Optional	Case study submitted demonstrating impact of change as a result of the programme	Green = 2
		Case study started or planned	Amber = 1
		Case study not started or planned	Red = 0
Compliance review via online platform (the portfolio review)	Optional	Case study submitted demonstrating impact of change as a result of the programme	Green = 2
		Case study started or planned	Amber = 1
		Case study not started or planned	Red = 0
Possible total score optional components			**12**
Possible total score core and optional components			**28**

The scores—total scores for core components, optional components, and both combined—are summarised. Complier Average Causal Effect (CACE) analyses were performed to account for compliance and engagement of the nurseries with the programme. An instrumental variable, two-stage least squares (2SLS) approach was used with random group allocation as the instrumental variable (Dunn, Maracy and Tomenson, 2005) and cluster standard errors to account for clustering at the nursery level.

The EEF analysis guidance suggests that the compliance indicator can be either binary or continuous and that 'minimal' and 'optimal' compliance thresholds can be defined and used to estimate bounds for the treatment effects. This pragmatic approach recognises that although a binary definition of compliance may be clear in some trials, in others it is far more challenging to identify the dividing line. However, the bounds used to define compliance can affect the likelihood that the exclusion restriction (a non-verifiable assumption of CACE analysis that the offer of the intervention should not have an effect on the outcome of those who do not comply with it) is violated. In general, there is a trade-off: less stringent definitions may underestimate the effect of engagement and make the analysis more similar to the ITT analysis but make the exclusion assumption more tenable while a more stringent definition of engagement may lead to larger CACE estimates but also in the exclusion restriction assumption being less realistic (Connell, 2009).

This is why recommendations are often to define compliance in multiple ways with varying levels of engagement and compare the results, which was the approach taken here.

Three CACE analyses for the primary outcome were conducted: one used the continuous compliance score considering the total score out of 16 for all the core components and two analyses defined compliance at the nursery level as a dichotomous variable:

- *nurseries that engaged at least minimally with the programme* (defined as the nursery being rated an amber score of 1 or a green score of 2 in at least one of the core aspects of the programme and a total core component score of at least 1 out of 16) as against *nurseries that received no intervention at all* (control nurseries plus all intervention nurseries for which all core components of the programme were rated red, score of 0); and
- *nurseries that delivered the programme with good fidelity* (defined as the nursery being rated an amber score of 1 or a green score of 2 in all of the core aspects of the programme—minimum score of 8 and all components scoring at least 1) as against *nurseries that delivered no intervention or delivered with poor fidelity* (control nurseries plus all intervention nurseries for whom at least one core component of the programme is rated red score of 0).

Results for the first stage (of the 2SLS process) are reported alongside (i) the correlation between the instrument and the endogenous variable (presented as the partial R^2 statistic from the first-stage estimation) and (ii) an F test (F statistic and p-value). The F statistic should exceed ten for inference based on the 2SLS estimator to be reliable when there is one endogenous regressor, as in this case (Bound, Jaeger and Baker, 1995; Stock, Wright and Yogo, 2002).

Additional analyses and robustness checks

In a sensitivity analysis for the primary outcome, we adjusted the primary analysis model for whether the child was tested at baseline by an assessor from the evaluation team or a nursery practitioner or teacher by including a child-level indicator for type of assessor as a fixed-effect covariate, plus an interaction of this factor with trial arm, to account for any hypothetical differences caused by type of assessor.

A further sensitivity analysis was conducted in which the primary analysis included an indicator variable for whether or not the post-test ASPECTS was conducted by a blinded RA, plus an interaction term with this factor and trial arm.

A similar sensitivity analysis was planned to account for the location of the post-test ('original' or 'new')—for children who left their original nursery and were assessed in an alternative location—however, all children were followed up in their original nursery so this analysis was not required.

Missing data analysis

The amount of missing primary baseline and post-test outcome data is summarised and reasons for missing data were explored and are provided in the report, where available. A multilevel logistic regression was used to model presence or absence of the primary outcome including all available child- and nursery-level baseline data as fixed effects, and nursery as a random effect. Significant predictors and possible mechanisms for the missing data are discussed.

The impact of missing data on the primary analysis was also assessed using multilevel imputation via the 'jomo' package in R (Quartagno and Carpenter, 2023).[4] Pre- and post-intervention ASPECTS maths score data was predicted by a linear regression model that includes all available child- and nursery-level baseline variables. This imputation procedure can account for the two-level (child and nursery) nature of the data.

A 'burn-in' of ten was used (meaning that the first ten iterations were discarded to allow the iterations to converge to the stationary distribution before the imputation) and 30 imputed datasets were created. The primary analysis was then rerun within the imputed datasets and Rubin's rules (Rubin, 1987) were used to combine the multiply-imputed estimates.

[4] The published SAP indicated that this analysis would be conducted via the REALCOM-impute macro, which is compatible with Stata (http://www.bristol.ac.uk/cmm/software/realcom/imputation.html) but this this software could not be obtained.

Secondary analysis

Language

The language score from ASPECTS was assessed at baseline and post-test. This is scored from 0 to 53, where a higher score indicates greater attainment. The language score from ASPECTS was analysed in the same way as the primary outcome, by comparing the score between the intervention and control groups using a linear mixed model at the child level. Group allocation, baseline ASPECTS language score, and nursery-level minimisation factors were included as fixed effects in the model and nursery as a random effect.

Practitioner confidence

Practitioner confidence (in teaching children maths) was assessed at post-test using a short online survey adapted from Chen et al. (2014). We requested for the survey to be completed by the nominated MC and DMC in intervention nurseries and comparable staff in control nurseries.

The 11-item 'confidence in helping nursery aged children learn maths' subscale was used. The MC or DMC, or equivalent practitioners, were asked to rate their agreement with each item on a Likert scale, from 1, 'strongly disagree', to 5, 'strongly agree'. Scores for items were summed to produce a summary score ranging from 11 to 55, with a higher score indicating greater confidence. The survey developers offer no guidance on how to handle missing item-level data for this instrument and so the scale was only scored if a valid response was provided across all 11 items.

Responses to items in the practitioner confidence survey are summarised descriptively by trial arm. These are presented for all respondents and disaggregated by MC and DMC of each nursery (where these persons can be identified).

The subscale score was compared between the two arms using a linear mixed model, adjusting for the nursery-level minimisation factors (number of graduate staff, nursery type, and nursery size) and highest qualification in maths of the respondent as fixed effects and nursery as a random effect.

Estimation of effect sizes

Effect sizes (ESs) were calculated by dividing the adjusted mean difference between the intervention and control group (accounting for prior attainment and the minimisation factors) by the pooled unconditional standard deviation obtained from the model run without these covariates. A 95% CI for the effect size was calculated by dividing the 95% confidence limits for the adjusted mean difference by this same denominator. All parameters used in these calculations are reported.

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

where, $(\overline{\mathrm{Y}}_{\mathrm{T}} - \overline{\mathrm{Y}}_{\mathrm{C}})_{\mathrm{adjusted}}$ denotes the adjusted difference in means between trial groups from the multilevel analysis model and sd_{pooled} denotes the pooled, unconditional standard deviation of the two groups (square root of the sum of the within- and between-cluster variances) from the unconditional model.

Implementation and process evaluation design

The implementation and process evaluation adhered to the key principles for the design, conduct, and reporting of IPEs (Humphrey et al., 2016). It was designed to be descriptive (cross-sectional and longitudinal), and to be delivered alongside and to complement, the impact evaluation. The IPE aimed to gain an understanding of usual practice (maths) in all participating nurseries and gather the perceptions and experiences of the MC programme from key stakeholders within intervention nurseries. The design of the IPE was refined following the results of the pilot trial (Appendix E).

Research methods

Data was gathered at baseline (pre-randomisation) as well as during and after the intervention delivery period. The IPE employed different methods to gather data relating to the pre-specified research questions and used triangulation to explore the relationship between intervention delivery and how it impacted on children's attainment. The nature of the data derived from the various data sources was descriptive: experiences and perceptions of the MC programme can be

inferred from this kind of data. Table 6 summarises the data collection methods used to address the IPE research questions.

Nurseries receiving the intervention were either randomly selected by trial statisticians or identified by the NDNA as 'best practice' nurseries (nurseries that engaged particularly well with the programme) and were invited by email to participate in semi-structured interviews held over the online video conferencing software, Zoom. The invitation included an interview-specific participant information sheet and a link to complete an eConsent form (using Qualtrics survey software). A pre-agreed, semi-structured interview schedule was implemented to assure the same questions were asked to all interview participants while allowing the opportunity for the interviewers to probe interesting lines of enquiry. In some cases, staff opted to participate in a focus group rather than individual interviews. Interviews and focus groups lasted between 30 and 60 minutes and were audio-recorded for transcription purposes only. Interviews and focus groups were undertaken cross-sectionally towards the end of the implementation period or after completion of the programme by two authors. One author undertook interviews or focus groups with participants from 13 nurseries and another with participants from one nursery. In addition, for quality assurance (QA) purposes, one author observed both of the other author's interviews and her focus group with one nursery and one author observed two of the other author's interviews. A semi-structured endpoint interview was also undertaken with the NDNA team.

Relevant staff within each participating nursery were requested to complete a survey at baseline (prior to randomisation) and post-intervention. All surveys were delivered online via Qualtrics survey software. At baseline, two surveys were issued to all nurseries to be completed by:

- the nominated MC within the nursery—which explored nominated staff qualifications, recently completed continuing professional development (CPD), reflection on maths teaching practice, and the child recruitment process; and
- the manager or headteacher—which explored staff composition, previous and planned nursery-level CPD in maths and other subjects, nursery budget and time committed to CPD, and the child recruitment process.

A further two surveys were issued post-delivery to all nurseries to be completed by:

- the nominated MC—exploring their qualifications, confidence in teaching maths, any CPD completed over the course of the academic year (excluding the MC programme), reflections on their own maths practice, and the effect of COVID-19 on the delivery of the MC programme (intervention nurseries only); and
- the manager or headteacher—exploring nursery-level CPD undertaken throughout the course of the academic year (excluding the MC programme), nursery budget and time committed to CPD, and the impact of COVID-19 on delivering the MC programme (intervention nurseries only).

The Early Childhood Environment Rates Scales-III (ECERS-3) and the ECERS-E, which has a specific focus on maths quality provision, were completed at baseline and post-intervention in a small subsample of nurseries (n = 3). ECERS data was collected by an external provider, Inquisitive Minds Matter Ltd. The purpose of collecting ECERS data was to provide insight and clarity to the results of the impact evaluation and on the impact of the MC programme on the quality of maths provision within nurseries.

E-log data summarising attendance at induction and subsequent training, as well as engagement with one to one support provided by the NDNA, was utilised to triangulate findings from other IPE data collection methods.

A focus group was conducted over Zoom with the delivery team lasting around 60 minutes.

Analysis

Quantitative survey data was summarised descriptively by randomised group. Continuous measures are reported as a mean and standard deviation while categorical data is reported as a count and percentage. No formal statistical comparisons were undertaken.

Analysis of the interview and focus group data was undertaken thematically and focused on RQ1, RQ2, RQ3, part of RQ4, and part of RQ7. Analysis was undertaken systematically with each of the 14 nurseries analysed sequentially until dominant views and experiences were clear, and until a range of relevant views were obtained where the experiences and perceptions diverged. RQ4 and RQ7 interview and focus group data was supplemented with attrition data from the

impact evaluation and data from Q27 in the two endpoint surveys (headteachers and nursery managers endpoint and Q10 (MCs and DMCs endpoint). We have included the dominant views of the participants to draw inferences, but we have also included a range of views, where appropriate, to reflect the diversity of practice experience and stakeholder perceptions.

Analysis of the survey data for RQ7 on the perceived impact of the COVID-19 pandemic was undertaken by looking at how frequently intervention nurseries reported the implementation of the programme had been affected by the pandemic. For those nurseries that reported an impact, we used thematic analysis to describe the ways in which the impact had been experienced and perceived. Themes originated from the research questions or emerged during the reading of the interview and focus group transcripts. Summaries within themes were made for each individual nursery; all nursery summaries were then combined to address each RQ focusing on dominant views and a range of views with an indication of the strength of evidence for each theme in terms of number of nurseries that expressed each perspective in each theme.

The remaining qualitative survey data was analysed using directed content analysis. Each survey was imported into Nvivo (v12), and, for each research question, the researcher read through the answers to the corresponding questions to gain insight into participants' responses. Following this, within each question, similar responses were grouped together, and different codes assigned to different groupings. The content within these codes was then summarised and used to answer the research questions (Hsieh and Shannon, 2005).

For each research question we analysed the extent to which the results support or contradict the relevant component of the logic model.

Table 6: IPE design and methods of data collection and analysis overview

Research methods	Data collection methods	Participants/data sources (type, number)	Data analysis methods	Research questions addressed	Implementation/ logic model relevance
Cross-sectional design	Semi-structured Interview/focus group	NDNA staff (n = 3)	Combination of inductive and deductive analysis with analyses grouped thematically according to RQs	RQ 1: 1.1; RQ 3: 3.4; RQ 4: 4.1; RQ 7	Fidelity; context
Cross-sectional design	Semi-structured Interview/focus group	MCs (n = 16)		RQ 1: 1.2, 1.3, 1.4; RQ 2: 2.1, 2.2, 2.3, 2.4; RQ 3: 3.1, 3.2, 3.4; RQ 7	Fidelity; process outcomes (confidence and competence)
Cross-sectional design	Semi-structured Interview/focus group	DMCs (n = 12)		RQ 1: 1.2, 1.3, 1.4; RQ 2: 2.1, 2.2, 2.3, 2.4; RQ 3: 3.1, 3.2, 3.4; RQ 7	Fidelity; process outcomes (confidence and competence)
Cross-sectional design	Semi-structured Interview/focus group	Other practitioners (n = 4)		RQ 2: 2.2, 2.3, 2.4; RQ 3: 3.1, 3.2, 3.3, 3.4; RQ 7	Fidelity; process outcomes (confidence and competence)
Longitudinal design	Log data of E-learning module attendance	All intervention settings	Frequency, counts; Regression	RQ 1: 1.1	Compliance; context
Longitudinal design	Nursery practice observation (ECERS-3 and ECERS-E)	PVI, n = 1; SN, n = 2	Descriptive analysis	RQ 3: 3.5	Context; outcomes

Longitudinal design	Baseline and endpoint nursery usual practice surveys	All control and intervention nurseries	Frequency; counts; descriptive and thematic analysis	RQ.5; 6; 7	Context

Costs evaluation design

The cost analyses followed the 'ingredients method' (Levin et al., 2017) to account for the costs of implementation at key stages.

Sources of data

It was originally intended that MCs and DMCs would complete three cost surveys (December 2021 and January 2022, February and March 2022, and June and July 2022), however, due to disruption to staffing within nurseries caused by the (omicron variant) COVID-19 outbreak between November 2021 and February 2022, which delayed some nurseries completing the programme start-up phase, we combined the relevant questions from the first two surveys (and pushed required completion back to March and April 2022 as opposed to February and March 2022 as specified in the protocol) to reduce overall burden on staff. Cost data was collected from participating nurseries via cost-specific surveys. MCs and DMCs were requested to complete two short, cost-specific, online surveys. A summary of the content of these surveys is provided below.

March and April 2022

> This survey captured the amount of time—staff working hours, paid and unpaid—spent completing the relevant training components of the programme (for example, the online induction, three two-hour e-learning training courses, use of audit tools, development of nursery-specific action plan, and one to one support received from the NDNA during the programme start-up phase). If staff indicated they had spent time completing start-up activities outside of normal working hours, the survey captured whether this was due to staff shortages caused specifically by COVID-19 or due to staff vacancies within their nursery (or both). These questions were added in response to a subsequent significant outbreak of COVID-19 in England between November 2021 and February 2022 and also high staff turnover within the early years sector, both of which coincided with the programme set-up. The survey also gathered data on nurseries' use of supply staff, and the cost where applicable, to cover key staff to complete any of the start-up activities associated with the programme and requested information on any start-up, prerequisite costs (for example, computer or internet connectivity) or unexpected or hidden costs associated with training.

June and July 2022

> These surveys captured the amount of time—staff working hours, paid and unpaid—involved in continuing to deliver the programme (for example, viewing monthly webinars, accessing resources via the MC online platform, monitoring and reviewing audit tools and the nursery's specific action plan, and one to one support from the NDNA), whether paid cover was arranged for staff to complete these activities, any recurring implementation costs (for example, materials, print outs, or resources), the use of existing resources, and unexpected or hidden costs.

At the end of intervention delivery, in June and July 2022, managers or headteachers were also requested to provide the full hourly cost (including wages, national insurance payments, benefits, and the cost of recruiting new teachers, among others) for each relevant staff member, for example, the MC or DMC.

Cost data was also collected directly from the delivery team in order to estimate the cost of delivering the training and support package.

Analysis

Time costs for nursery personnel are summarised descriptively. The cost of training was calculated by summing the cost of the constituent parts to estimate a total. This was then divided by the total number of randomised intervention nurseries (66) to estimate a per nursery cost. The cost of the optional extras, staff cover, and materials was calculated

by summing each item to estimate a total; this was then divided by the total number of randomised intervention nurseries to estimate a per nursery cost.

Per-child costs were determined by summing the total costs per nursery and dividing by the average number of children per nursery eligible for inclusion in this evaluation (34).

Assumptions

- For the purpose of calculating the costs over a three year period, we assumed that the costs of additional resources captured in the final survey would be recurring. This is based on the assumption that some materials may need to be replaced annually or additional, similarly-priced materials may need to be purchased. Materials costs are considered as optional extras given the inconsistent uptake by nurseries.
- In the first survey, nurseries were asked to specify the actual costs of providing cover. In the follow-up survey, nurseries were only asked to provide the number of hours cover required. In order to calculate the cost of cover, we used the hourly rates provided by the nursery manager or headteacher and assumed that the hourly cost of covering the MC or DMC was equivalent to their respective average hourly rate of pay. This assumes an equivalently qualified and paid individual provided cover.

Timeline

Table 7: Timeline

Dates	Activity	Staff responsible or leading
12 Jul 2019	Set-up meeting 1	EEF, evaluation team (ET), delivery team (DT)
2 Sep 2019	Set-up meeting 2	EEF, ET, DT
10 Oct 2019	IDEAs workshop	ET, DT
Sep–Nov 2019	Protocol development	ET
Nov 2019	Ethics application for pilot study	ET
Dec 2019–Jan 2020	Ethics application for effectiveness trial	ET
Nov 2019–Feb 2021	Pilot study	DT, ET
Jan–Oct 2021	Recruit nurseries	DT (support from ET)
July–Nov 2021	Recruit parent/carers; schedule baseline assessments; baseline assessment training	ET
2 Sep 2021	Autumn term begins	-
Oct–Dec 2021	Baseline assessments with children; nursery usual practice survey	ET
Oct–Dec 2021	Batch randomisation	ET
Nov–Dec 2021	ECERS baseline within identified IPE intervention nurseries	Inquisitive Minds Matter Ltd (ECERS only, external providers)
25–29 Oct 2021	School half-term	-
Mid Oct 2021–Jun 2022	Delivery of MC programme (support and resources provided for 7–8 months)	DT
Jan 2022	SAP submitted	ET

Dates	Activity	Staff responsible or leading
Sep 2021–Aug 2022	IPE interviews	ET
Jun–Jul 2022	Post-test outcome assessments with children; practitioner confidence and beliefs survey; ECERS post-intervention within identified IPE intervention nurseries	ET and Inquisitive Minds Matter Ltd (ECERS only, external providers)
23 Jul–31 Aug 2022	School summer holidays	-
Aug 2022	IPE DT interview	ET
Sep–Dec 2022	Confirmation of 'school destination of children' collected via nurseries/parents/carers to enable matching to National Pupil Database.	ET
Sep–Dec 2022	Data analysis and report writing	ET
Dec 2022	Submit impact and IPE draft report for pilot and effectiveness trial	ET
Apr 2023	Submission of final edited EEF report, submission of data to the EEF data archive and updating the ISRCTN trial registry with results; submission of interim statement of spend to date.	ET
Long term follow-up		
Nov 2023	Submission of National Pupil Database request for Early Years Foundation Stage Profile data (completed at the end of reception)	ET
Nov 2023–Jan 2024	Report addendum analysis and writing	ET
Feb 2024	Submit addendum long-term follow-up	ET
Apr 2024	Submission of long-term data to the EEF archive and updating of ISRCTN trial registry with results; submission of final statement of spend to the EEF.	ET

Impact evaluation results

Participant flow including attrition and exclusions

Nursery recruitment and attrition

Due to the recruitment methods used, such as advertising on the NDNA website, the exact number of nurseries that were approached to participate in the evaluation is not known. A total of 602 completed an expression of interest survey on that website. When nursery eligibility criteria were applied against these survey responses, 402 (66.8%) nurseries initially appeared to meet the eligibility criteria while 200 nurseries did not meet them all. Potentially eligible nurseries were contacted by the NDNA to confirm eligibility and were requested to complete and return a MoU if they wanted to take part in the trial, as well as a DSA (to cover data sharing with the University of York) and an EULA (required to use CEM's online ASPECTS assessment). As part of the eligibility criteria for participation, nurseries had to agree to all requirements outlined in each of these three documents.

In total, 153 nurseries agreed to participate by returning a fully completed MoU, DSA, and EULA. Of these, seven did not progress to the child recruitment phase (that is, were not asked to distribute information to parents or carers) due to the following reasons:

- one nursery was excluded due to being unresponsive after three contact attempts to confirm its eligibility;
- one that was on the reserve list no longer wanted to take part when offered a place in the trial in late September 2021; and
- five were considered ineligible: one due to being in the same nursery chain or academy trust as another participating nursery (only one nursery per nursery chain or academy trust could participate, as per the eligibility criteria); and one due to reporting on the MoU that they only had 12 eligible children (a minimum of 15 was an eligibility criterion).

While nurseries were only sent the MoU if they reported on the expression of interest survey that they would have at least 15 eligible children during the 2021/2022 academic year, nurseries were only able to provide an estimate at that stage of the year. Three nurseries that went on to be randomised into the trial estimated on the MoU that they would have 14 eligible children and one nursery that was randomised estimated 13 eligible children. These nurseries reported being confident that they could recruit at least ten eligible children and were ultimately brought on board in the interests of meeting the recruitment targets (the nursery with 13 eligible children was the final reserve nursery to start the trial, as other reserves were prioritised).

In total, 146 nurseries were asked to start child recruitment. During the child recruitment phase and before nursery randomisation, 12 were excluded:

- eight due to staffing issues or capacity constraints, with three specifically mentioning COVID-19 or sick leave among staff as a barrier; and
- four due to collecting too few consent forms (nurseries were only randomised if six or more consent forms were returned): two had exhausted recruitment and two had become unresponsive to contact from the evaluation team.

It total, 134 nurseries were randomised: 66 to the intervention group and 68 to the control. Following randomisation, there were no withdrawals from the evaluation prior to post-testing. At the time of post-testing, all 134 nurseries remained in the evaluation.

Four intervention nurseries withdrew from the MC programme (but were retained for evaluation) due to staffing and capacity issues.

Child recruitment and attrition

The figures reported in this section apply to nurseries that went on to be randomised and excludes those that were withdrawn pre-randomisation.

On the MoU, nurseries provided the estimated number of children within their nursery who would be eligible to participate in the pre- and post-intervention ASPECTS assessments—at three to four years of age—due to begin reception in September 2022 and attending the nursery for a minimum of 15 hours per week. As child recruitment began prior to September 2021, nurseries provided an estimate of the number of eligible children they would have in their 2021/2022 cohort. The (estimated) total number of children across participating nurseries identified as eligible was 4,611 (per nursery: mean, 34.4; SD, 19.6; median, 30; range, 13 to 132). All 134 nurseries that went on to be randomised provided data on child eligibility.

Consent forms for 1,954 children were returned to the evaluation team. Of these, 126 (6.4%) children were excluded due to the following reasons:

- 96 children were excluded due to not meeting the child eligibility criteria or meeting the exclusion criteria; and
- 30 children were excluded due to invalid consent (i.e., the consent form was not signed or all consent statements were not initialled/ticked).

In total, parent or carer consent was gained for 1,828 eligible children, 739 via paper consent and 1,089 via e-consent. The eligible consenting children were then randomly ordered within nurseries by one of the trial statisticians for selection to complete the baseline assessment with the aim of obtaining assessments for an average of ten children per nursery.

Baseline assessments using ASPECTS

ASPECTS was completed with 1,304 children at baseline (intervention, *n* = 638 (48.9%); control, *n* = 666 (51.1%); per nursery: mean, 9.7; SD, 0.8; median, 10; range, 6 to 11).

Of the children who completed the baseline assessment, 1,301 had been randomly selected for baseline assessment as per the protocol, however, three from three different nurseries had been selected by their nursery (all three were later post-tested and included in the primary analysis). In these three cases, one of the ten pre-selected children was absent during the assessment period so the nursery self-selected a reserve child to assess (that is, an eligible child with parent/carer consent who had not initially been selected for assessment by the evaluation team) instead of contacting the evaluation team to select the next available child from the randomly ordered list of reserve children.

Eight nurseries reported that they did not have staff capacity to assess the children at baseline using ASPECTS and agreed to a visit by a trained member of the evaluation team to assess the children. In total, 52 children (4.0%) were assessed by a member of the evaluation team at baseline. The remaining 1,252 were assessed by a practitioner within the nursery, as per protocol. Nurseries were not randomised and informed of their allocation until they had completed the baseline testing.

Post-testing

Post-testing was only supposed to be conducted with children who had completed the baseline assessment, therefore, post-testing was expected to include 1,304 children. There was a mix-up at one nursery and one child who had not completed baseline tests was post-tested in error (this child was not included in the randomised sample, in the summaries below, or in the analyses). The post test figures are as follows:

- total number of post-tests completed: *n* = 1,209/1,304 (92.7%); intervention, *n* = 600/638 (94.0%); control, *n* = 609/666 (91.4%); per nursery: mean, 9.0; SD, 1.2; median, 9; range, 5 to 11;
- missing post-test data: *n* = 95/1,304 (7.3%); intervention, n = 38/638 (6.0%); control, n = 57/666 (8.6%);
    - left—informed before post-testing: *n* = 55; intervention, *n* = 17; control, *n* = 38; and
    - left—informed during post-testing: *n* = 8; intervention, *n* = 7; control, *n* = 1.

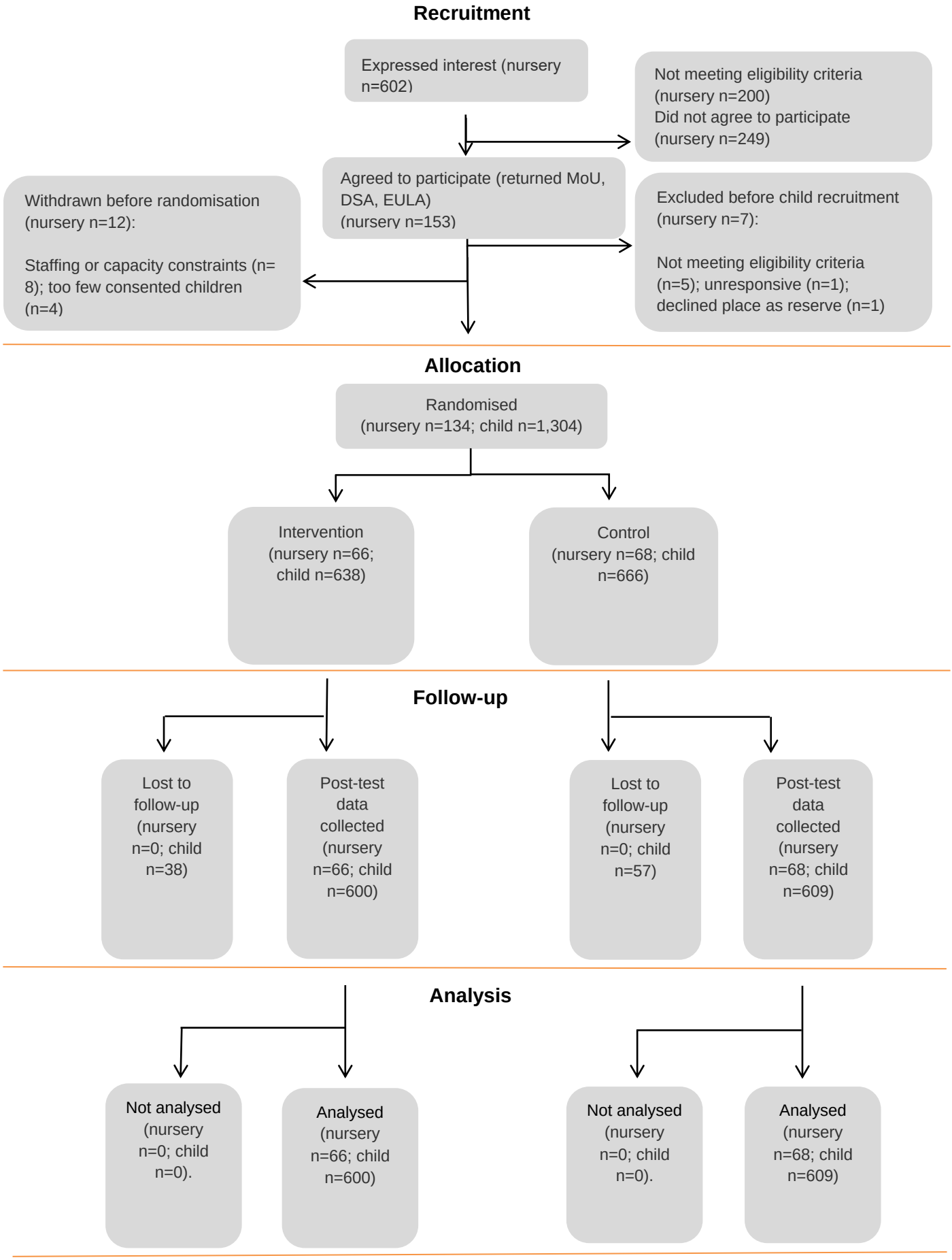
      [Of the 694 children from randomised PVIs, 39 (5.6%) left the nursery before post-testing; this was lower in randomised SN or maintained nurseries at 24 (3.9%) of the 610 children who had left].
    - absent on assessment days: *n* = 25; intervention, *n* = 10; control, *n* = 15;
    - child refused: *n* = 6; intervention, *n* = 3; control, *n* = 3; and
    - nursery declined—child with SEN: *n* = 1 (intervention).

Post-tests were conducted with 1,176 children by 11 trained research assistants, blinded to random allocation, however some children were absent on the days or times when RAs were scheduled to visit. Where feasible, RAs made a second or third visit to nurseries to try to assess all randomised children. Where a blinded RA was not available (for example, due to illness or rails strikes), or in cases where it was not feasible to send another RA (for example, if the nursery was remote or if the remaining children to assess had poor attendance), practitioners at the nursery were requested to complete the ASPECTS post-test with the remaining children. In total, 26 (2.2%) children (intervention, *n* = 16; control *n* = 10) across 21 nurseries (intervention, *n* = 16; control, *n* = 9) were assessed at post-test by a practitioner at the nursery, and seven (0.6%) children (intervention, *n* = 0; control, *n* = 7) from one (control) nursery were assessed by a member of the evaluation team.[5]

shows a flow diagram of nurseries and participants through the trial. A total of 1,209 children of the 1,304 enrolled were included in the primary analysis; this equates to an attrition rate of 7.3%.

[5] This occurred as the RA scheduled to complete the assessments was delayed due to a rail strike and a member of the evaluation team was available at the nursery as they were due to complete a quality assurance observation of the assessments.

Figure 2: Participant flow diagram



As stated in the protocol, we aimed to have 80% power to detect an effect size of 0.20 of a standard deviation with 1,380 children, assuming a baseline/-post-test correlation of 0.59, an ICC of 0.17, and 15% loss to follow-up (Table 4). At randomisation, we anticipated having baseline data for 1,304 children and so, under otherwise identical assumptions, calculated the minimum detectable effect size (MDES) to be unchanged (0.20). The actual observed ICC at the nursery level obtained from the primary analysis model was 0.16 (95% CI: 0.11 to 0.21). The overall correlation between the pre- and post-test scores of the participants included in the primary analysis was 0.61. Based on the number of children included in the primary analysis model (n = 1,209), and the observed ICC and baseline/post-test correlation, the estimated MDES at analysis for the primary outcome was 0.19 (Table 8).

Table 8: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	EYPP*	Overall	EYPP*	Overall	EYPP**
MDES* **		0.20	0.38 / 0.30	0.20	0.30 / 0.39	0.19	0.36
Baseline/ post-test correlations	Level 1 (child)	0.59	0.59	0.59	0.59	0.61	0.70
	Level 2 (nursery)	N/A	N/A	N/A	N/A	N/A	N/A
Intracluster correlations (ICCs)	Level 2 (nursery)	0.17	N/A / 0.17	0.17	N/A / 0.17	0.16	0.02
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two	Two	Two
Average cluster size		10	1 / 2	9.7	1 / 2	9	2.4
Number of nurseries	Intervention	69	69	66	66	66	27
	Control	69	69	68	68	68	36
	Total:	138	138	134	134	134	63
Number of children	Intervention	690	69 / 138	638	66 / 132	600	66
	Control	690	69 / 138	666	68 / 136	609	88
	Total:	1,380	138 / 276	1,304	134 / 268	1,209	154

EYPP = Early Years Pupil Premium.
* Figures either side of the / represent the two scenarios: (i) aggregating data to nursery-level and (ii) conducting analysis at child level.
** Child-level.
*** Estimates at protocol and randomisation stages assume 15% child-level attrition at post-test.

Attrition

Overall, 1,209 (92.7%) of the 1,304 randomised children had a valid baseline and post-test ASPECTS score and so were included in the primary outcome analysis model. Across trial groups, the ratios (analysed to randomised) were 600:638 in the intervention arm and 609:666 for the control arm (Table 9).

Table 9: Child-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of children	Randomised	638	666	1,304

	Analysed	600	609	1,209
Child attrition— from randomisation to analysis	Number	38	57	95
	Percentage	6.0	8.6	7.3

Child and nursery characteristics

Characteristics for the 134 randomised nurseries and 1,304 participating children are presented in Table 10 and appear broadly similar between the two groups. The number of children from each nursery ranged from six to 11 (median ten) in the intervention group, and seven to 11 (median ten) in the control group. No formal hypothesis testing was performed on baseline data (Senn, 1994), so comparisons were made visually only.

Nurseries in the intervention group had, on average, 2.2 graduate staff members; 61 (92.4%) nurseries had at least one graduate staff member. Similarly, the control group had, on average, 2.4 graduate staff members per nursery, with 63 (92.6%) having at least one graduate staff member.

There were 35 (53.0%) PVI nurseries in the intervention group and 36 (52.9%) in the control group, and the rest were maintained or SN. The majority were private or independent—32 (48.5%) in the intervention group and 32 (47.1%) in the control group—and a small number were voluntary: three (4.5%) in the intervention group and four (5.9%) in the control group.

Approximately half of the children were male: 316 (49.5%) in the intervention group and 330 (49.5%) in the control group. Regarding EYPP eligibility status among children, there are 50 (7.8%) missing datapoints in the intervention group and 44 (6.6%) in the control group. For children with EYPP data, 72 (12.2%) in the intervention group were eligible for EYPP and 99 (15.9%) in the control group were eligible for EYPP.

Some differences were observed for English as an Additional Language (EAL) between groups. There is a larger proportion of missing data in the control group compared to the intervention group: 70 (10.5%) versus 32 (5.0%) missing data points in the intervention group and 70 (10.5%) in the control group. For children with EAL data, the control group had a larger proportion of EAL children; in the intervention group there were 53 (8.7%) children with EAL and in the control group there was 83 (13.9%). There is also a larger proportion of children eligible for FEEE at two years old (defined by the composite FEEE variable) in the control group (38.9%) than in the intervention group (32.0%).

Participating children attended nursery for an average of 24.6 (SD, 8.7) hours per week in the intervention group and 24.2 (SD, 9.3) in the control.

At the time of writing, there are no sources of data we can identify that report on this same data at the national level, which would allow for a comparison between our recruited sample of nurseries and children and the national picture to assess representativeness. The best available source is a government report published annually on education provision for children under five, the latest publication being June 2022.[6] This contains statistics on the early years provision for three- and four-year-olds, and eligible disadvantaged two-year-olds, who are entitled to 570 hours annually of government-funded early years provision. In 2022, of all providers (excluding childminders) delivering the 15- or 30-hour entitlement to three- and four-year-olds, 62.5% were PVI providers and 37.5% were state-funded schools. In this trial, just over half (52.9%) of the recruited nurseries were a PVI setting and 47.1% a SN or maintained setting.

Baseline ASPECTS data was collected from all 1,304 children across 134 nurseries (average, 9.7 per nursery; SD, 0.8; range 6 to 11). The average score for the maths component was 12.6 (SD, 6.9) out of a possible 29 in the intervention group and 12.8 (SD, 6.6) in the control group (Hedges' g effect size between the groups 0.04; 95% CI: -0.07 to 0.15). The average score for the language component was 24.4 (SD, 9.1) out of a possible 53 in the intervention group and 24.6 (SD, 8.3) in the control group (Hedges' g effect size between the groups 0.02; 95% CI: -0.08 to 0.13).

[6] https://explore-education-statistics.service.gov.uk/find-statistics/education-provision-children-under-5

Characteristics for the 1,209 children included in the primary analysis are presented in Table 11. Given all nurseries were included in the primary analysis model, nursery-level data is unchanged. Child-level characteristics and baseline outcome scores were very similar between the ‘as randomised’ and ‘as analysed’ samples.

Table 10: Baseline characteristics of groups as randomised

Nursery level (categorical)	**Intervention group**		**Control group**	
	n/N (missing)	**Count (%)**	**n/N (missing)**	**Count (%)**
>1 graduate staff	66/66 (0)	61 (92.4)	68/68 (0)	63 (92.6)
Nursery type PVI SN and Maintained	66/66 (0)	 35 (53.0) 31 (47.0)	68/68 (0)	 36 (52.9) 32 (47.1)
Nursery type subgroups Private/independent School-based Maintained Voluntary	66/66 (0)	 32 (48.5) 21 (31.8) 10 (15.2) 3 (4.5)	68/68 (0)	 32 (47.1) 20 (29.4) 12 (17.6) 4 (5.9)
Nursery level (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**
Number of graduate staff	66/66 (0)	2.2 (1.6)	68/68 (0)	2.4 (1.9)
Number of eligible children	66/66 (0)	35.4 (22.4)	68/68 (0)	33.5 (16.6)
Child-level (categorical)	**n/N (missing)**	**Count (%)**	**n/N (missing)**	**Count (%)**
Sex, male	638/638 (0)	316 (49.5)	666/666 (0)	330 (49.5)
Eligible for EYPP	588/638 (50)	72 (12.2)	622/666 (44)	99 (15.9)
EAL	606/638 (32)	53 (8.7)	596/666 (70)	83 (13.9)
Eligible for FEEE at 2 years old (parent/carer response)	571/638 (67)	167 (29.2)	594/666 (72)	212 (35.7)
Received FEEE at 2 years old (parent/carer response)	629/638 (9)	184 (29.3)	661/666 (5)	239 (36.2)
Received FEEE at 2 years old (nursery response)	575/638 (63)	119 (20.7)	553/666 (113)	145 (26.2)
Eligible for FEEE at 2 years old (composite of above 3 variables)	638/638 (0)	204 (32.0)	666/666 (0)	259 (38.9)
EYFS2Y personal, social, emotional development [a] 0–20m 16–26m 22–36m 30–60m	300/638 (338)	 12 (4.0) 61 (20.3) 187 (62.3) 40 (13.3)	272/666 (394)	 12 (4.4) 55 (20.2) 163 (59.9) 42 (15.4)

EYFS2Y Communication and Language [a] 0–20m 16–26m 22–36m 30–60m	302/638 (336)	 15 (5.0) 69 (22.8) 175 (57.9) 43 (14.2)	273/666 (393)	 12 (4.4) 63 (23.1) 149 (54.6) 49 (17.9)	
EYFS2Y Physical development [a, b] 0–20m 16–26m 22–36m 30–50m 40–60m	300/638 (338)	 <10 (-) 63 (21.0) 185 (61.7) 45 (15.0) <10 (-)	273/666 (393)	 <10 (-) 42 (15.4) 165 (60.4) 55 (20.1) <10 (-)	
Child level (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**Hedges' *g* effect size (95% CI)**
Hours attendance at nursery per week	638/638 (0)	24.6 (8.7)	666/666 (0)	24.2 (9.3)	N/A
ASPECTS maths score	638/638 (0)	12.6 (6.9)	666/666 (0)	12.8 (6.6)	0.04 (-0.07, 0.15)
ASPECTS language score	638/638 (0)	24.4 (9.1)	666/666 (0)	24.6 (8.3)	0.02 (-0.08, 0.13)

[a] The age ranges (in months) refer to developmental age of the child at the time the EYFS2Y check, as reported by the completing practitioner.
[b] Cell counts less than ten suppressed to prevent statistical disclosure.

Table 11: Comparison of baseline child characteristics, as included in the primary analysis

	Intervention group		**Control group**		
Child level (categorical)	**n/N (missing)**	**Count (%)**	**n/N (missing)**	**Count (%)**	
Sex, male	600/600 (0)	299 (49.8)	609/609 (0)	299 (49.1)	
Eligible for EYPP	543/600 (57)	66 (12.2)	507/609 (102)	88 (17.4)	
EAL	572/600 (28)	48 (8.4)	548/609 (61)	73 (13.3)	
Eligible for FEEE at 2 years old (parent/carer response)	536/600 (64)	151 (28.2)	541/609 (68)	188 (34.8)	
Received FEEE at 2 years old (parent/carer response)	591/600 (9)	169 (28.6)	604/609 (5)	212 (35.1)	
Received FEEE at 2 years old (nursery response)	543/600 (57)	106 (19.5)	507/609 (102)	125 (24.7)	
Eligible for FEEE at 2 years old (composite of above 3 variables)	600/600 (0)	188 (31.3)	609/609 (0)	229 (37.6)	
EYFS2Y Personal, social, emotional development[a] 0–20m 16–26m 22–36m 30–60m	276/600 (324)	 12 (4.3) 55 (19.9) 177 (64.1) 32 (11.6)	245/609 (364)	 11 (4.5) 44 (18.0) 150 (61.2) 40 (16.3)	

EYFS2Y Communication and Language[a] 0–20m 16–26m 22–36m 30–60m	278/600 (322)	 15 (5.4) 61 (21.9) 166 (59.7) 36 (12.9)	246/609 (363)	 11 (4.5) 48 (19.5) 140 (56.9) 47 (19.1)	
EYFS2Y physical development [a, b] 0–20m 16–26m 22–36m 30–50m 40–60m	276/600 (324)	 <10 (-) 57 (20.7) 174 (63.0) 38 (13.8) <10 (-)	246/609 (363)	 <10 (-) 36 (14.6) 147 (59.8) 53 (21.5) <10 (-)	
Child-level (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**Hedges' g effect size (95% CI)**
Hours attendance at nursery per week	600/600 (0)	24.6 (8.8)	609/609 (0)	24.5 (9.3)	N/A
ASPECTS maths score	600/600 (0)	12.6 (6.9)	609/609 (0)	13.0 (6.7)	0.05 (-0.07, 0.16)
ASPECTS language score	600/600 (0)	24.5 (9.0)	609/609 (0)	24.9 (8.3)	0.05 (-0.06, 0.16)

[a] The age ranges (in months) refer to developmental age of the child at the time the EYFS2Y check, as reported by the completing practitioner.
[b] Cell counts less than ten suppressed to prevent statistical disclosure.

Outcomes and analysis

Primary analysis

A valid baseline ASPECTS maths score was obtained from all 1,304 randomised children from 134 nurseries (100%: intervention, n = 638; control, n = 666). The ICC (95% CI) associated with nursery for the baseline score is 0.18 (0.13 to 0.25). In total, a valid post-test ASPECTS maths score was available for 1,209 children (92.7%: intervention, n = 600, 94%; control, n = 609, 91.4%) from 134 nurseries (intervention, 66; control, 68). A mean of 17.9 (95% CI: 17.4 to 18.4) was observed in the intervention arm and 16.5 (95% CI: 15.9 to 17.0) in the control arm. The unadjusted mean difference is 1.43 (95% CI: 0.72 to 2.14) (Appendix D, Appendix Table 3). Histograms of the pre- and post-test scores show they are roughly normally distributed (Figure 3) and were available for 1,209 (92.7%) children (intervention, n = 600, 46%; control, n = 609, 46.7%). The correlation between the pre- and post-test scores was 0.61 (95% CI: 0.57 to 0.64). As a check of the analysis model assumptions, the normality of the standardised residuals was checked using a QQ plot and were shown to be normal (Figure 4). The adjusted mean difference in post-test score between the intervention and control groups was 1.58 (95% CI: 0.75 to 2.42, p < 0.001, Appendix D, Appendix Table 3). The estimated Hedges' g effect size was 0.25 (95% CI: 0.12 to 0.38), which relates to approximately three months' additional progress in the intervention group (Table 12). The total variance used to calculate the effect size (from a model without covariates) was 39.77—the sum of 34.41 (random variation between children, within-cluster variance) and 5.36 (heterogeneity between nurseries, between-cluster variance). The ICC associated with nursery from the adjusted model was 0.16 (95% CI: 0.11 to 0.21). The ICC for the empty model (that is, without covariates) was 0.13 (95% CI: 0.09 to 0.19).

Figure 3: Histogram of pre- and post-intervention ASPECTS maths scores for the randomised sample (values at the extremes of the range have been grouped together to avoid statistical disclosure where there were values with fewer than ten pupils).

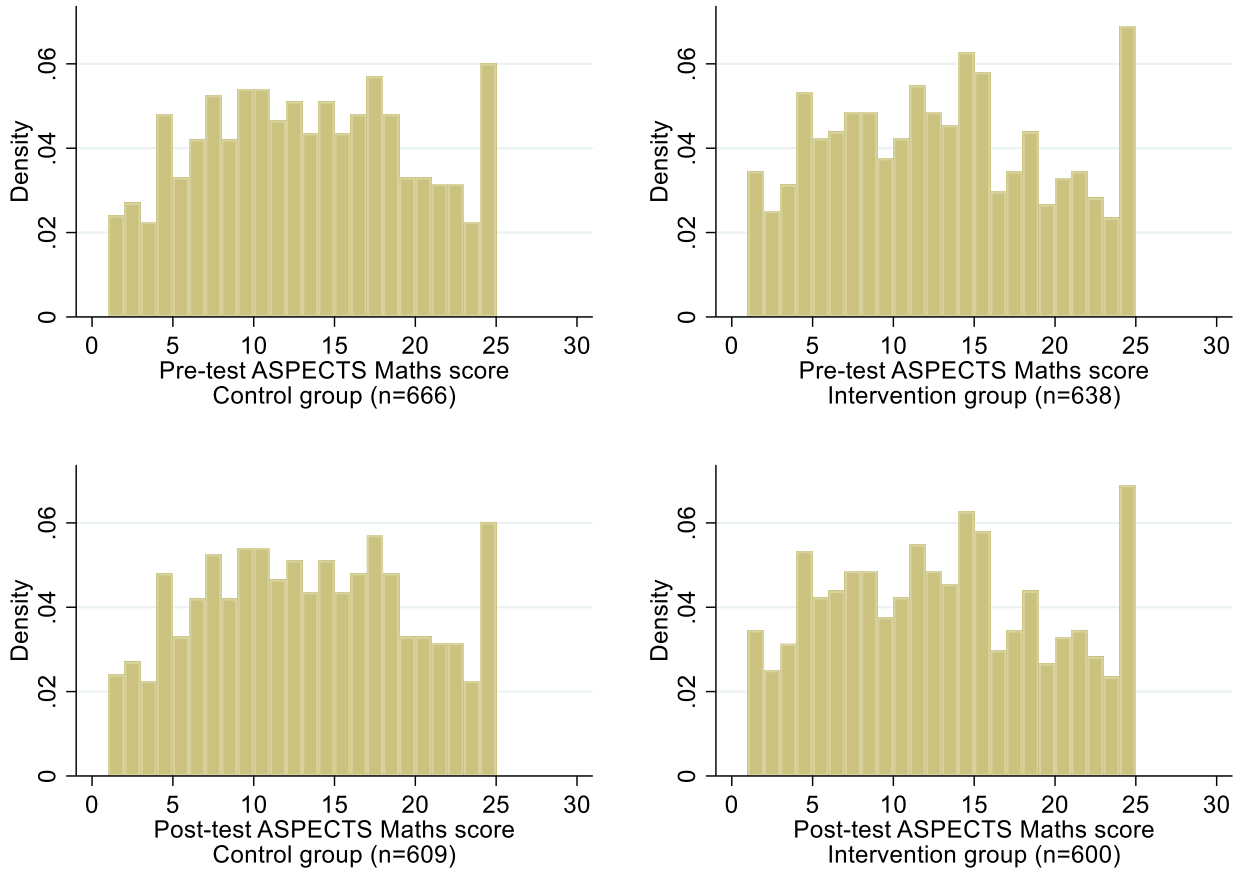
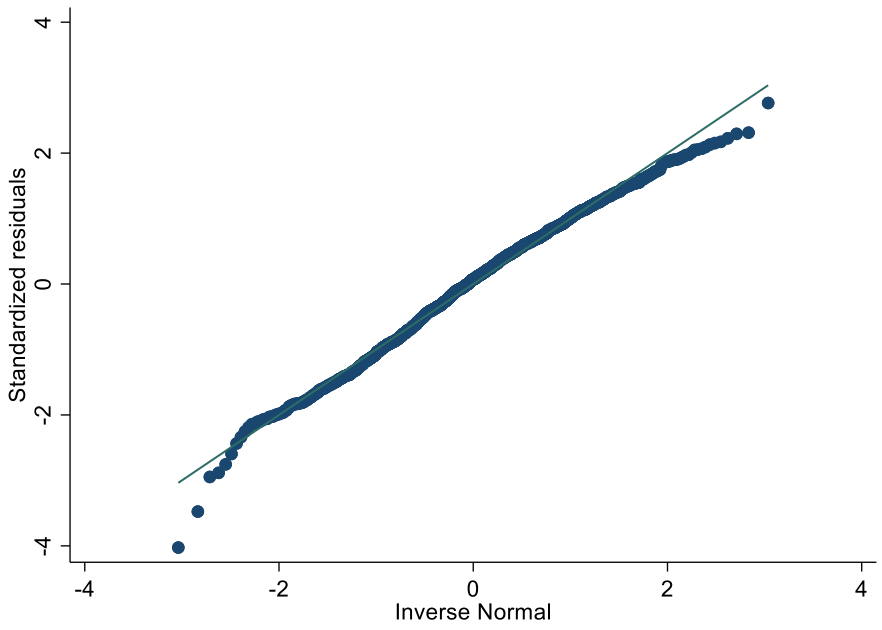


Figure 4: QQ plot of the standardised residuals from the primary analysis model to assess the normality assumption[7]



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

Table 12: Primary analysis

	Unadjusted means (post-intervention)				Effect size		
	Intervention		Control				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges g (95% CI)**	**p-value**
Post-test ASPECTS maths score	600 (38)	17.9 (17.4, 18.4)	609 (57)	16.5 (15.9, 17.0)	1,209 (600; 609)	0.25 (0.12, 0.38)	<0.001

Subgroup analyses

Summary statistics for the post-test ASPECTS maths score are presented in Table 13 by gender, average number of hours the child attends the nursery per week (dichotomized at the median of 24 hours), eligibility for EYPP, eligibility for FEEE at two years old, and nursery type. These summaries indicate that, in general, children attending nursery for more than 24 hours per week performed slightly better on the post-test than those who attended nursery less than this. Scores for children who were eligible for EYPP, or eligible for FEEE at two years old, tended to be lower than for those who were not. Scores were similar for male and female children, and for children attending PVI compared with SN.

In a series of adjusted regression analyses that included interaction effects, there was no evidence of an interaction between trial allocation and (i) gender (interaction term coefficient 0.59, 95% CI: -0.44 to 1.62, p = 0.26), (ii) average number of hours child attends nursery per week (dichotomised at 24 hours, 0.01, 95% CI: -1.14 to 1.17, p = 0.98), (iii) eligibility for FEEE (0.58, 95% CI: -0.57 to 1.74, p = 0.32), or (iv) type of nursery child attended (PVI vs SN; -0.59, 95% CI: -2.30 to 1.11, p = 0.49). There was, however, some evidence of an interaction between trial allocation and eligibility for EYPP (interaction term coefficient 1.65, 95% CI: -0.10 to 3.41, p = 0.07). The adjusted mean difference between intervention and control in the EYPP subset from the interaction effects model was 3.11 (95% CI: 1.47 to 4.74).

There were 154 children from 63 nurseries who were eligible for EYPP and had a valid baseline and post-test ASPECTS maths score (median two per nursery, range one to seven in both groups). Across these 63 nurseries, 23 (36.5%) only had one participating child who was eligible for EYPP, which meant that 63.5% of these nurseries had more than one child eligible for EYPP. Within the restricted sample of children eligible for EYPP, the correlation between the baseline and post-test scores was 0.63 (95% CI: 0.53 to 0.72). The adjusted mean difference in post-test score between the intervention and control groups within this sample was 3.08 (95% CI: 1.33 to 4.82 p < 0.001, Appendix D, Appendix Table 3) and the estimated Hedges' g effect size was 0.47 (95% CI: 0.20 to 0.73), which relates to approximately six months' additional progress in the intervention group. The total variance used to calculate the effect size (from a model without covariates) was 43.44—the sum of 37.45 (random variation between children, within-cluster variance) and 5.99 (heterogeneity between nurseries, between-cluster variance). The ICC associated with nursery from the adjusted model was 0.02 (95% CI: 0.00 to 1.00) and from the empty model (that is, without covariates) was 0.14 (95% CI: 0.03 to 0.48).

Table 13: Subgroup summary scores for the post-test ASPECTS maths score

	Raw means			
	Intervention		Control	
Outcome	**n**	**Mean (SD)**	**N**	**Mean (SD)**
Gender				
Male	299	17.9 (6.2)	299	16.6 (6.5)
Female	301	17.9 (5.9)	310	16.3 (6.5)
Number of hours at nursery				
< 24	282	17.5 (6.2)	300	15.9 (6.5)
	318	18.2 (5.9)	309	17.0 (6.4)

≥ 24				
Eligible for EYPP Yes No	66 488	17.0 (6.9) 18.2 (5.8)	88 482	12.9 (5.8) 17.1 (6.4)
Eligible for FEEE at 2 years old Yes No	188 412	16.7 (6.3) 18.4 (5.8)	229 380	14.2 (6.1) 17.8 (6.3)
Nursery type PVI SN	315 285	18.3 (5.9) 17.4 (6.1)	315 294	16.8 (6.3) 16.1 (6.6)

Analysis in the presence of non-compliance

Intervention nurseries were first contacted by the NDNA to begin the MC programme between 21 October 2021 and 7 March 2022. Four nurseries withdrew from delivering the programme an average of 2.4 months after first being contacted. The remaining 62 intervention nurseries delivered the programme. The average time between first NDNA contact and date of withdrawal or post-testing within the nursery was 7.1 months (SD 1.5, range 0–8.6).

All intervention nurseries (n = 66) were assessed by the NDNA for their fidelity to the intervention. All control nurseries were given a fidelity score of 0/16. RAG (red, amber, green) ratings for the eight core intervention components are summarised in

Table 14. For all eight components, over half of the nurseries were scored green (‘very engaged’). For the 66 intervention nurseries, the average score for the core components out of a possible 16 was 13.6 (SD: 3.0), median 15, range 4 to 16.

In the CACE analysis using the continuous compliance score (total score out of 16 for all the core components), the CACE estimate of the effect of engaging with the intervention on the children’s maths attainment was a predicted increase of 0.12 points (95% CI: 0.06 to 0.18, $p < 0.001$; effect size 0.02, 95% CI: 0.01 to 0.03). The partial R^2 from the first stage of the CACE analysis was 0.93, and the F-statistic was $F (1, 133 = 1,659.9, p < 0.001)$; these indicate a strong correlation between the instrumental variable (random allocation) and the endogenous variable and that the inference of the CACE estimate is reliable.

Of the 66 intervention nurseries, all 66 were defined as engaging at least minimally with the intervention. Therefore, the CACE analysis considering the effect of engaging at least minimally with the intervention on the children’s maths attainment was not necessary as this would be equivalent to the ITT analysis.

Of the 66 intervention nurseries, 49 (74.4%) were defined as being very or partially engaged in all of the core aspects of the intervention (that is, delivering the intervention with good fidelity). The CACE estimate of the effect of being very or partially engaged on the children’s maths attainment was a predicted increase of 2.13 points (95% CI: 0.99 to 3.27, $p < 0001$; effect size 0.34, 95% CI: 0.16 to 0.52). The partial R^2 from the first stage of the CACE analysis was 0.60 and the F-statistic was $F (1, 133 = 193.1, p < 0.001)$; these indicate reasonable correlation between the instrumental variable (random allocation) and the endogenous variable and that the inference of the CACE estimate is reliable.

In both CACE analyses, the effects suggest that children performed better when the nursery complied more fully with the intervention.

Table 14: Summary of fidelity RAG rating for core components of the MC programme as assessed by the NDNA.

Core components	RAG rating	Number (%) of intervention nurseries
Identification of suitable MC (MC; graduate or level 3 practitioner).	Green = 2	66 (100.0)
	Amber = 1	0 (0.0)
	Red = 0	0 (0.0)
Identification of suitable Deputy MC (DMC; qualified to at least level 3).	Green = 2	66 (100.0)
	Amber = 1	0 (0.0)
	Red = 0	0 (0.0)
MC and DMC complete induction.	Green=2	56 (84.8)
	Amber = 1	9 (13.6)
	Red = 0	1 (1.5)
Completion by the MC of 2 courses: ‘developing mathematical confidence in the early years: the big ideas of number sense’; ‘developing mathematical thinking in the early years: shape space, measures and pattern – including characteristics of effective learning and sustained, shared thinking’.	Green = 2	56 (84.8)
	Amber = 1	7 (10.6)
	Red = 0	3 (4.5)
Use of audit tool.	Green = 2	35 (53.0)
	Amber = 1	22 (33.3)
	Red = 0	9 (13.6)
Completion and continued use of an action plan.	Green = 2	36 (54.5)
	Amber = 1	18 (27.3)
	Red = 0	12 (18.2)
Use of up to 10 mandatory resources provided through online platform: 3- to 4-year-olds: ‘build a maze’, ‘number hunt’, ‘delivering the post’, ‘mud kitchen’, ‘cars down a ramp’, ‘patterns’, ‘construction’, ‘tidy up time’, ‘snack time’, ‘outdoor games’.	Green = 2	44 (66.7)
	Amber = 1	9 (13.6)
	Red = 0	13 (19.7)
Engagement with one to one support provided by the NDNA.	Green = 2	48 (72.7)
	Amber = 1	17 (25.8)
	Red = 0	1 (1.5)

For the 66 intervention nurseries, the average (SD) score for the optional components out of a possible 12 was 5.0 (2.8), median 5, range 0 to 11 (

Table 15).

Table 15: Summary of fidelity RAG rating for optional components of the MC programme as assessed by the NDNA.

Optional Components	RAG rating	Number (%) of intervention nurseries
Track and monitor development of 6 children on termly basis.	Green = 2	24 (36.4)
	Amber = 1	22 (33.3)
	Red = 0	20 (30.3)
Monthly webinars.	Green = 2	34 (51.5)
	Amber = 1	15 (22.7)
	Red = 0	17 (25.8)
Completion by the DMC 2 courses: 'developing mathematical confidence in the early years: the big ideas of number sense'; 'developing mathematical thinking in the early years: shape space, measures and pattern – including characteristics of effective learning and sustained, shared thinking'.	Green = 2	15 (22.7)
	Amber = 1	9 (13.6)
	Red = 0	42 (63.6)
Completion by MC/DMC of Coaching as an Educational Lead course.	Green = 2	7 (10.6)
	Amber = 1	34 (51.5)
	Red = 0	25 (37.9)
Reflection and completion of case study based on outcomes of action plan.	Green = 2	8 (12.1)
	Amber = 1	0 (0.0)
	Red = 0	58 (87.9)
Compliance review via online platform.	Green = 2	33 (50.0)
	Amber = 1	7 (10.6)
	Red = 0	26 (39.4)

Additional analyses and robustness checks

An assessor from the evaluation team conducted baseline assessments in eight nurseries (n = 52 children); the rest were assessed by a practitioner within the nursery. The intervention effect was little changed when the primary analysis model was adjusted for whether the child was tested at baseline by an assessor from the evaluation team or a practitioner/teacher in their nursery, this was included as a fixed-effect covariate, plus as an interaction with trial arm (adjusted mean difference 1.66, 95% CI: 0.80 to 2.52, $p < 0.001$; effect size 0.26, 95% CI: 0.13 to 0.40). The pooled total variance used to calculate the effect size was 39.77 as in the primary analysis.

A member of the evaluation team completed post-test outcome assessments with seven children in one nursery (this was during a QA visit as the assessor was late due to rail strikes). A practitioner/teacher completed ASPECTS in 21 nurseries with a total of 26 children. The intervention effect was also little changed when the primary analysis model was adjusted for whether or not the post-test ASPECTS was conducted by a blinded RA, this was included as a fixed-

effect covariate, plus as an interaction with trial arm (adjusted mean difference 1.54, 95% CI: 0.70 to 2.38, p < 0.001; effect size 0.24, 95% CI: 0.11 to 0.38). The pooled total variance used to calculate the effect size was 39.77 as in the primary analysis.

Missing data analysis

Overall, 95 of 1,304 (7.3%) children are excluded from the primary analysis due to missing post-test data (intervention, n = 38 of 638, 6.0%; control, n = 57 of 666, 8.6%; Table 9). Reasons for this were predominantly because the child had left the nursery or was absent on the day of testing, as described in the participant flow section.

In the adjusted mixed-effect logistic regression with presence or absence of post-test ASPECTS maths score as the outcome variable, two of the included covariates were observed to be statistically significant predictors, assessed at the 5% level. These were nursery type (children attending PVI nurseries were less likely to have missing data compared to SN and maintained nurseries) and EAL (children with EAL were less likely to have missing data). These variables were included in the multiple imputation. Missing post-test ASPECTS data was imputed using multiple imputation by chained equations. The primary analysis model was rerun on the multiply-imputed data set and Rubin's rules were used to combine the treatment estimates.

The adjusted mean difference in ASPECTS post-test maths score between the two groups following multiple imputation was 1.58 (95% CI: 0.72 to 2.45, p < 0.001). The estimated Hedges' g effect size was 0.25 (95% CI: 0.12 to 0.39), which relates to approximately three months' additional progress in the intervention group and is the same as the results of the primary analysis. The total variance used to calculate the effect size was 39.41—the sum of 34.42 (within-cluster variance) and 4.99 (between-cluster variance).

Secondary analysis

ASPECTS language score

A valid baseline ASPECTS language score was obtained from all 1,304 randomised children from 134 nurseries (100%: intervention, n = 638; control, n = 666). The ICC (95% CI) associated with nursery for the baseline score is 0.17 (0.12 to 0.23). In total, a valid post-test ASPECTS language score was available for 1,209 children (92.7%: intervention, n = 600, 94%; control, n = 609, 91.4%), from 134 nurseries (66 intervention, 68 control). A mean of 31.7 (95% CI: 30.9 to 32.5) was observed in the intervention arm and 30.0 (95% CI: 29.2 to 30.7) in the control arm. The unadjusted mean difference is 1.68 (95% CI: 0.60 to 2.76) (Appendix D, Appendix Table 3). Histograms of the baseline and post-test scores show they are roughly normally distributed (Figure 5). Baseline and post-test scores were available for 1,209 (92.7%) children (intervention, n = 600, 46%; control, n = 609, 46.7%). As a check of the analysis model assumptions, the normality of the standardised residuals was checked using a QQ plot and were shown to be normal (Figure 6). The correlation between the baseline and post-test scores was 0.61 (95% CI: 0.57 to 0.65). The adjusted mean difference in post-test score between the intervention and control groups was 2.06 (95% CI: 0.73 to 3.39, p = 0.002; Table 16). The estimated Hedges' g effect size was 0.21 (95% CI: 0.08 to 0.35), which relates to approximately three months' additional progress in the intervention group. The total variance used to calculate the effect size (from a model without covariates) was 92.29—the sum of 76.13 (random variation between children, within-cluster variance) and 16.16 (heterogeneity between nurseries, between-cluster variance). The ICC associated with nursery from the model was 0.18 (95% CI: 0.12 to 0.24). The ICC for the empty model (that is, without covariates) was 0.18 (95% CI: 0.13 to 0.24).

Figure 5: Histogram of pre- and post-intervention ASPECTS language scores for the randomised sample (values at the extremes of the range have been grouped together to avoid statistical disclosure where there were values with fewer than ten pupils).

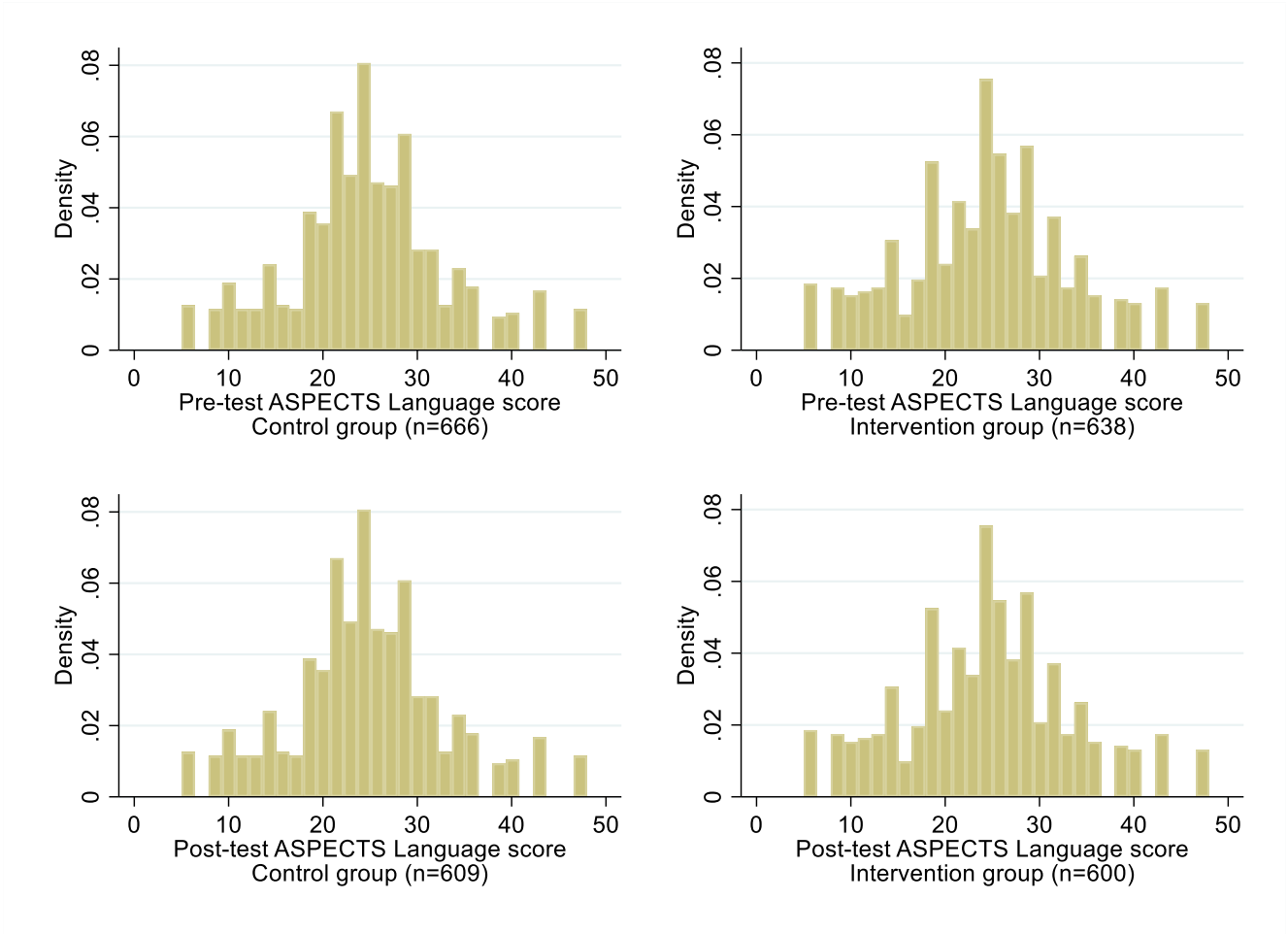
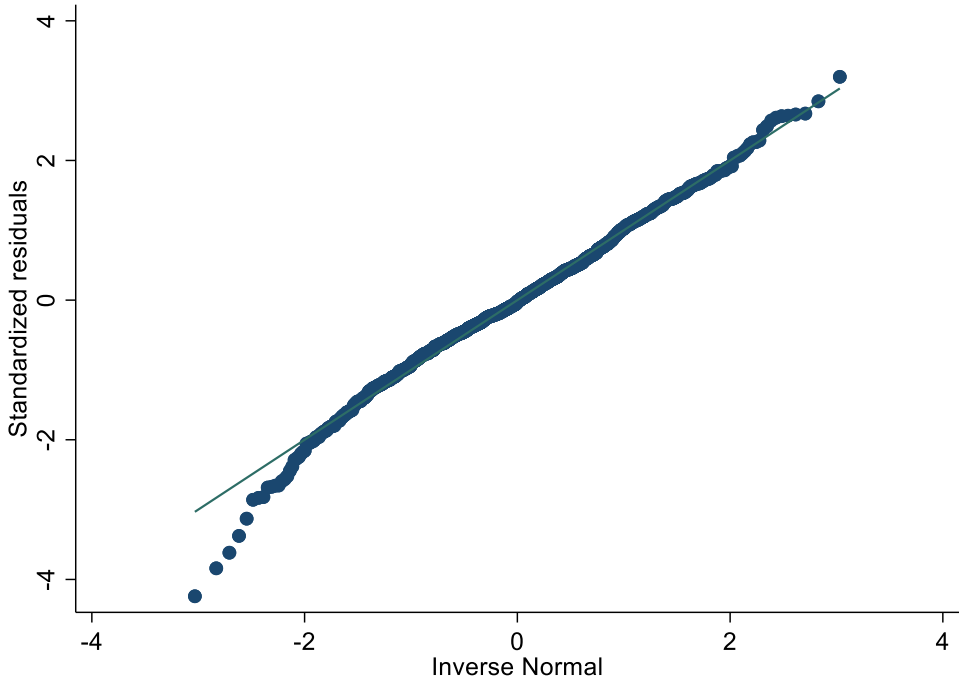


Figure 6: QQ plot of the standardised residuals from the secondary analysis model to assess the normality assumption



Practitioner confidence survey

A total of 139 practitioner confidence surveys (intervention group, n = 75; control group, n = 64) were received from 93 nurseries (49 in the intervention group and 44 in the control group). A response was therefore received from 69.4% of the randomised nurseries. A link to the online practitioner confidence survey was emailed to nurseries in July 2022, and two reminders were subsequently issued to encourage practitioners to complete the survey. Given that this survey was issued at the end of the trial, staff changes since the start of the trial may have impacted the response rate (that is, nurseries may not have had a nominated MC or DMC by summer 2022). A response from the MC of a nursery could be identified for 37 (56.1%) of the 66 intervention nurseries. Responses to the individual items are presented in Technical Notes Appendix 7 for all respondents by randomised group and for the MCs and DMCs separately.

Some evidence of a difference was observed between the intervention and control groups in the scores for subscale two ('confidence in helping nursery aged children learn maths'). The adjusted mean difference between the intervention and control groups for subscale two was 2.35 (95% CI: -0.12 to 4.81, p = 0.06); the effect size was calculated as 0.35 (95% CI: -0.02 to 0.71; Table 16). The total variance used to calculate this effect size (from a model without covariates) was 45.69—the sum of 43.78 (random variation between children, within-cluster variance) and 1.91 (heterogeneity between nurseries, between-cluster variance).

Table 16: Secondary analysis

	Raw 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**
Post-test ASPECTS language	600 (38)	31.7 (30.9, 32.5)	609 (57)	30.0 (29,2, 30.7)	1,209 (600; 609)	0.21 (0.08, 0.35)	0.002
Confidence in helping nursery aged children learn maths	75 (0)	48.1 (46.8, 49.4)	64 (0)	46.0 (44.1, 48.0)	139 (75; 64)	0.35 (-0.02, 0.71)	0.06

Feasibility of accessing ASQ-3 data completed when children were two years old

Alongside the evaluation of the MC programme, we continued our exploratory work—started within the pilot study but hampered due to COVID-19—on the potential use of the ASQ-3 for early years research trials (Appendix E). Since 2015, the ASQ-3 has been used as part of the Healthy Child Programme health and developmental review that all children in England should receive between two and two and a half years of age (Department of Health and Social Care, 2016; Public Health England, 2018; 2020). We intended to explore the feasibility of accessing routinely collected ASQ-3 data, as well as the coverage of this data.

Furthermore, if feasible and if coverage was sufficient, we aimed to access ASQ-3 data for children participating in this trial and explore how well ASQ-3 scores at age two correlate with maths and language development at three to four years (that is, post-test ASPECTS maths and language scores). This would inform researchers on the potential strategy of linking to routinely collected ASQ-3 data for use as a baseline measure.

Health professionals such as health visitors help parents to complete the 24-, 27-, or 30-month version of the ASQ-3, as appropriate, and data is required to be submitted to NHS Digital, specifically the Community Services Dataset (Public Health England, 2018). We contacted the NHS Digital enquiries team who advised that data linkage to routinely collected ASQ-3 data may be possible and such requests would need to be made through the Data Access Request Service. Through completing the Health Research Authority NHS Research Ethics Committee (REC) decision tool, it was determined that NHS REC approval would be required to link to routinely collected ASQ-3 data, which can be a time-consuming process.

The NHS Digital noted that official statistics for ASQ-3 were still being published by Public Health England and NHS Digital would take over reporting of this data when their data quality is high enough.

There is limited published information on coverage estimates of ASQ-3 data from the two- to two-and-a-half-year check held by NHS Digital. Between October 2016 and March 2017, ASQ-3 data was recorded within NHS Digital for just 12% of the eligible children over this period (NHS Digital and Ofsted, 2017). Public Health England (recently replaced by the Office for Health Improvement and Disparities) has been collecting and reporting ASQ-3 data via submissions from local authorities on an interim basis until the Community Services Dataset, operated by NHS Digital, take on reporting of this. Recent guidance states that this interim reporting arrangement is still in place as data submitted through the Community Services Dataset does not have sufficient coverage for reporting purposes (Public Health England, 2021; Office for Health Improvement and Disparities, 2022). Due to this insufficient coverage, it was decided it was not worth pursing linking to ASQ-3 data held by NHS Digital.

In contrast to NHS Digital, Public Health England has better coverage of ASQ-3 data obtained using the interim reporting system. The cohort of children eligible for the current trial would have been due for a two- to two-and-a-half-year check between 1 September 2019 and 1 March 2021. In 2019/2020, 78.6% of children in England received this check by age two and a half and 92.6% of these had ASQ-3 completed (Public Health England, 2022). The corresponding figures for 2020/2021 were 71.5% and 85.2% respectively (Office for Health Improvement and Disparities, 2021). While Public Health England has good coverage rates, even during COVID-19, exploring the feasibility of linking to ASQ-3 data held by Public Health England would not be useful for future research since the data reporting is migrating to NHS Digital.

As part of the consent form for the current effectiveness trial, parents and carers agreed for their child's nursery to provide ASQ-3 data, completed for their child at two years old, if held by the nursery. We explored how many participating nurseries held ASQ-3 data during our phone calls to nurseries in spring 2022, which were primarily carried out to establish whether participating children still attended the nursery (that is, to monitor attrition). The first 25 nurseries phoned were asked whether they had completed the ASQ-3 with the current cohort of three- to four-year-old children at age two and whether they held this data. Five said 'yes' (20%) and 20 said 'no' (80%), indicating that the majority of nurseries do not hold this data. This limited exercise suggested that accessing ASQ-3 data from nurseries for use as a baseline measure in future early years trials is not feasible as a requirement to hold and provide this data would be a barrier to nursery recruitment. Mindful of the need to minimise the burden on nurseries, and to focus our time and resources on key tasks for the evaluation of the MC programme, further nurseries were not asked if they held ASQ-3 data nor was it requested from nurseries.

In summary, accessing ASQ-3 data via NHS Digital was ultimately determined to be infeasible due to insufficient coverage of ASQ-3 data within NHS Digital's Community Services Dataset. In addition, accessing ASQ-3 data held by nurseries was also deemed unrealistic due to insufficient coverage in our sample and the decision to minimise burden on nurseries. Due to this, it was not feasible to access ASQ-3 data for children participating in this trial and explore the correlation between ASQ-3 scores at age two with ASPECTS maths and language scores at three to four years.

Implementation and process evaluation results

This section presents a summary of the data collected and analysed as part of the implementation and process evaluation. The IPE was cross-sectional and longitudinal and comprised data collected across the evaluation. A full overview of the IPE design and research questions was presented earlier in Table 6.

Summary of IPE data collected by method

Interviews and focus groups

It was originally planned that relevant staff members from 11 nurseries (two SN, nine PVI) would be recruited to participate in Zoom interviews or focus groups (whichever was most convenient for the participating nursery). A total of 30 nurseries were invited to participate in these: 20 from the intervention group in general (ten of which were chosen at random by trial statisticians and a further ten chosen non-specifically—the next ten on the list) plus ten chosen non-specifically (the first ten on the list) from the group of nurseries that the NDNA identified as examples of 'best practice'.

In total, 33 staff members across 14 nurseries (six SN/maintained and eight PVI) provided relevant consent and participated in one to one interviews or focus groups. Participants included 16 MCs, 12 DMCs, and a further four staff members who did not have a role associated with delivering the MC programme. Of the 14 nurseries that agreed to participate, ten were rated by the NDNA (as part of compliance data) as being 'very engaged' (eight PVI, two SN), three as being 'partially engaged' (two PVI, one SN), and one as 'not engaged' (SN). Of these 14 participating nurseries, seven were identified by the NDNA as examples of 'best practice'.

The total number of Intervention nurseries was 66 so 14 represents over 20%. Table 17 provides a comparison of nursery-level characterises between the group of nurseries that participated in the IPE and all nurseries randomised to the intervention group. Nurseries in the IPE sample had fewer graduate staff in comparison to nurseries in the randomised intervention group. Despite this, the IPE sample is sufficiently large to enable some generalisability from the sample to the population (all intervention nurseries), although please refer to the Limitations and Lessons Learnt section within the Conclusion section for further commentary on nursery sampling for interviews and focus groups.

Table 17: Comparison of characteristics between randomised intervention nurseries and IPE nurseries

Nursery level (categorical)	**Intervention (randomised, n = 66)**		**Intervention (IPE sample, n = 14)**	
	n/N (missing)	**Count (%)**	**n/N (missing)**	**Count (%)**
>1 graduate staff	66/66 (0)	61 (92.4)	14/14 (0)	12 (85.7)
Nursery type PVI SN and Maintained	66/66 (0)	 35 (53.0) 31 (47.0)	14/14 (0)	 5 (35.7) 9 (64.3)
Nursery type subgroups Private/independent School-based Maintained Voluntary	66/66 (0)	 32 (48.5) 21 (31.8) 10 (15.2) 3 (4.5)	14/14 (0)	 3 (21.4) 5 (35.7) 4 (28.6) 2 (14.3)
Nursery level (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**
Number of graduate staff	66/66 (0)	2.2 (1.6)	14/14 (0)	1.8 (1.3)

Number of eligible children	66/66 (0)	35.4 (22.4)	14/14 (0)	31.1 (26.5)

The NDNA participated in an online focus group following intervention delivery that documented its experience of delivering the MC programme.

Baseline and endpoint surveys

Baseline surveys

The baseline surveys were sent to all participating nurseries between October and November 2021. One survey was to be completed by the headteacher or manager and another was to be completed by the nominated MC (prior to randomisation, all nurseries were required to identify a nominated MC who would receive the intervention should the nursery be allocated at random to receive it and be a point of contact within the nursery throughout the trial period).

In total, 131 headteachers and managers responded to the baseline survey. Of these, 20 responses were invalid (responders did not answer any questions), and one was from a nursery that withdrew prior to randomisation, leaving 110 responses from 94 nurseries (14 nurseries had two responses each and one nursery had three). Since the questions ask about the nursery, duplicate responses were removed and the first complete response from each nursery was selected for analysis. Therefore, 94 responses were analysed (intervention, 48; control, 46).

There were a total of 141 responses to the nominated MC baseline survey. Of these, 23 were invalid (responders did not answer any questions), and one was from a nursery that withdrew prior to randomisation, leaving 117 responses from 108 nurseries (nine had two responses each, which were included in the analyses). Therefore, 117 responses were analysed (intervention, 57; control, 60).

Endpoint surveys

The endpoint surveys were sent to all participating nurseries in July 2022 and two reminders were sent to encourage completion. One survey was to be completed by the headteacher or manager and another by the nominated MC.

In total, 99 headteachers and managers responded: of these, four responses were invalid (responders did not answer any questions) leaving 95 responses from 83 nurseries (ten nurseries had two responses each and one had three responses). Since the questions ask about the nursery, duplicate responses were removed, and the first complete response from each nursery was selected for analysis. Therefore, 83 responses were analysed (intervention, 41; control, 42). In total, 90 MCs responded to the endpoint survey. Of these, five responses were invalid (responders did not answer any questions) leaving 85 responses from 76 nurseries (seven nurseries had two responses each and one had three responses). Therefore, 85 responses were analysed (intervention, 41; control, 44).

Table 18 provides an overview of the number of valid survey responses by allocation and staff role within the evaluation.

Table 18: Survey data (pre and post) total number of valid responses

Intended respondent	Control		Intervention	
	Baseline survey (n)	Endpoint survey (n)	Baseline survey (n)	Endpoint survey (n)
Headteacher/manager	46	42	48	41
Nominated MC	60	44	57	41

Early Childhood Environment Rates Scales-III

As noted within the protocol, for the IPE we aimed to recruit four intervention nurseries to receive a visit from an independent researcher at Inquisitive Minds Matter Ltd to complete ECERS-3 and the ECERS-E. In total, three nurseries provided relevant consent and completed ECERS-3 and ECERS-E data. These nurseries were also invited to participate in IPE interviews and focus groups, however only one nursery agreed for staff to be interviewed. The baseline visits

occurred following randomisation in November and December 2021 and follow-up visits occurred in June and July 2022. Inquisitive Minds Matter Ltd provided nurseries with ECERS feedback at the end of the trial in August 2022. It was not possible to recruit the target number of nurseries (four) as during this phase of the trial nurseries began restricting entry to non-essential visitors due to the omicron COVID-19 wave. The purpose of collecting ECERS data was to provide insight and clarity to the results of the impact evaluation and the impact, if any, of the MC programme on the quality of maths provision within nurseries.

Usual practice

This section explores the IPE results in relation to usual practice for early maths and maths CPD across all participating nurseries prior to randomisation (baseline) and explores what usual practice was for control nurseries during the trial. The IPE research questions addressed within this section are:

RQ5 What is 'usual practice' in all nurseries?

RQ6 What maths-related professional development (PD) opportunities do staff have in control group nurseries?

6.1. What are the perceived impacts of these maths-related PD opportunities on nursery staff's maths-related classroom practice?

6.2. What are the perceived impacts of these maths-related PD opportunities on children's maths attainment?

6.3. What other maths-related PD opportunities are nursery staff looking for?

Data from the baseline and endpoint surveys informs this section.

Usual practice in all nurseries

The baseline survey asked nominated MCs, 'What does usual maths teaching/provision look like for three- to four-year olds within your nursery?' . Here, the majority of survey responses indicated that maths was part of everyday practice and was delivered as continuous provision. Commonly reported activities included identifying the day of the week, number of the week, talking about shapes and colours, counting steps as they climb, playing number games, counting at registration, setting up the correct number of chairs, exploring vocabulary that relates to size and shape, pattern recognition through puzzles, and counting how many children were in the class that day. Most nurseries did a daily ten-minute small group activity that was maths based. Maths was also introduced in key play times through the strategies listed above and staff used mathematical language in everyday practice. Nurseries also stated that maths talk and practical skills are incorporated into everything they did and were used to lead in to every learning session.

Nominated MCs were also asked what tools they used to monitor children's progress: by far the most commonly used was 'Tapestry', which was used to record observations and track progress. Tracking sheets to document a child's progress and development were also commonly used. This was often used with notes against each child to show their next steps and challenges. Photographic and video recordings as well as written and verbal observations were used. End of term reports and parent discussions and meetings were other tools used to monitor a child's progress. Development Matters Checkpoints, Tiny Tracker, EYFS tracker Opal Assessment, *Evidence Me* (online learning journal), Online Learning Profile, Early Essence, O Track, Insight, the 'Birth to 5 Matters' document, and Family app were other tools used by nurseries to monitor and record progress.

Nominated MCs were asked to provide examples of learning activities given to children to complete at home. Going on nature walks was a very common home activity, this often involved collecting particular items on the walk and bringing them in to the nursery to discuss. Children were also asked to look for numbers and shapes in the environment. Helping to set the table and counting the cutlery or helping to do the shopping and counting the items needed were also well utilised home activities. Another common home activity was cooking and baking and asking the children to pick the right ingredients and look at the weights required. A number of nurseries created their own 'home learning' bags which were sent home monthly with children on rotation. Sometimes maths challenges were given out to be completed over holidays. Parents reading with children was often suggested as a home activity and could also include a weekly reading book, a song of the week, or a number of the week and number songs. Name-writing and other counting activities were also encouraged at home, such as sorting out the socks in the washing basket, counting the fruit in the fruit bowl, playing catch and counting each time you catch, and sorting and counting toys when tidying.

As detailed in Table 19, intervention nurseries reported having a greater number of rooms or classes of three- to four-year-olds in comparison to the control group, however the number of staff dedicated to those rooms was similar between randomised groups. There was also a similar number of graduate level staff within intervention and control nurseries.

Table 19: Summary of headteacher and manager responses to baseline survey by randomised group.

Variable (categorical)	Intervention group		Control group		All nurseries	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Role	48/48 (0)		46/46 (0)		94/94 (0)	
Nursery manager or deputy manager		32 (66.7)		26 (56.5)		58 (61.7)
Senior nursery manager, nursery director, or owner		0 (0)		3 (6.5)		3 (3.2)
Teacher or nursery teacher		2 (4.2)		1 (2.2)		3 (3.2)
EYFS lead		1 (2.1)		6 (13.0)		7 (7.4)
Headteacher or deputy headteacher		11 (22.9)		10 (21.7)		21 (22.3)
Other		2 (4.2)		0 (0)		2 (2.1)
Nursery has undertaken other maths/numeracy-related CPD programmes in last 2 years	47/48 (1)	8 (17.0)	46/46 (0)	13 (28.3)	93/94 (1)	21 (22.6)
Nursery has undertaken CPD programmes that are not maths/numeracy related in last 2 years	47/48 (1)	24 (51.1)	45/46 (1)	27 (60.0)	92/94 (2)	51 (55.4)
Nursery staff have a certain amount of allocated time each month/year for CPD	46/48 (2)	21 (45.7)	45/46 (1)	26 (57.8)	91/94 (3)	47 (51.6)
Nursery has other CPD planned for 2021/2022 (other than the Maths Champions programme)	46/48 (2)	23 (50.0)	45/46 (1)	32 (71.1)	91/94 (3)	55 (60.4)
Variable (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**
Number of rooms/classes of 3- to 4-year-olds	48/48 (0)	1.9 (3.7)	46/46 (0)	1.5 (0.8)	94/94 (0)	1.7 (2.7)
Number of teaching staff/practitioners based in 3- or 4-year-old classes	46/48 (2)	4.6 (2.2)	46/46 (0)	4.9 (2.5)	92/94 (2)	4.8 (2.4)
Number of nursery staff qualified to level 6 (graduate) or above in early years	48/48 (0)	2.3 (1.7)	46/46 (0)	2.0 (1.6)	94/94 (0)	2.1 (1.7)
Nursery's annual budget for CPD training (£)	23/48 (25)	1,456.5 (1,640.0)	26/46 (20)	1,673.1 (1,941.1)	49/94 (45)	1,571.4 (1,790.8)

Headteachers or managers were asked to indicate the nursery's annual CPD budget. Nurseries in the control group reported this to be £217 (13%) higher than intervention nurseries prior to the trial. Similarly, staff in control nurseries were more likely to be allocated time each month/year to undertake CPD in comparison to staff in intervention nurseries. This may go some way in explaining why a greater proportion of control nurseries reported having undertaken maths or numeracy CPD and CPD unrelated to maths or numeracy in the previous two years in comparison to the intervention group (see Table 18). There was a wide range of maths-related training that had been undertaken across various nurseries such as in-house, online, and face to face external workshops. Specific examples included NDNA online courses, Primary Advantage Maths INSET days, Maths Hub, online training regarding mathematical development and provision in early years, and the early years development maths strand. Some courses were completed through the Early Excellence Centre or the National Centre for Excellence in the Teaching of Mathematics. Some nurseries were provided training by the local authority. Additionally, a small proportion of nurseries (fewer than five) indicated that they were also participating in the Department for Education's Early Years Professional Development Programme concurrently with this trial, however the distribution of such nurseries was evenly balanced between control and intervention groups. As detailed in Table 18, prior to randomisation, more control nurseries (60%) reported implementing non-maths or numeracy-related CPD in the previous two years in comparison to the intervention group (51%). The types of training courses undertaken by all nurseries spanned speech, language and communication (for example, WellComm Early Years, Eklan, and I Can Talk), PSED (for example, Healthy Smiles Oral health training, Health Early Years, anxiety in children), and safeguarding.

As detailed in Table 18, prior to randomisation, more control nurseries (71%) planned to implement CPD during the year of the trial in comparison to those in the intervention group (50%). In terms of CPD to be implemented by nurseries in the near future, some nurseries said that training is decided nearer the time on a needs basis and others reported planning to implement programmes or CPD spanning variety of domains including language and literacy (for example, Jolly Phonics, Language Champion training, WellComm Early Years, Eklan, or Tales Toolkit), PSED (for example, oral health training or The Attachment and Trauma Sensitive Nursery Award), special educational needs and disabilities (for example, The Balanced System or Makaton), parental engagement (Parent Champion Training), safeguarding, and first aid (for example, Millie's Mark). A small number of nurseries were implementing forest schools or a Hygge approach. As demonstrated here, nurseries cited a variety of upcoming programmes or CPD initiatives which appears more expansive than those reported to have been undertaken in the previous two years. This could be a result of the pandemic and the limits it imposed on nurseries to undertake such training or other programmes and also the priorities of the nurseries during that time.

Usual practice in the control group

Within the endpoint survey, headteachers and managers in the control group were asked what, if any, early years maths- or numeracy-related CPD their staff had undertaken since the start of the trial. Of the 42 respondents, eight (19%) indicated that their staff had undertaken some form of maths/numeracy related CPD since September 2021: it included CPD from the maths lead and in house training, EYFS e-learning, and non-specified EEF maths courses over nine months. Of the 42 respondents in the control group, six indicated that since September 2021 they had participated in the Department for Education's Early Years Professional Development Programme. In total, 17/42 (41%) of control nurseries who responded to the endpoint survey reported that they had completed training unrelated to maths or numeracy since September 2021. Such CPD or training included safeguarding, first aid, courses around speech and language (communication and phonics), literacy, food hygiene, outdoor play, health and safety, and positive physical interaction and metacognition. Some responses indicated that the training covered many topics and was too broad to list. CPD in a weekly staff meeting was also common.

Headteachers and managers reported that staff in 20 of the 42 responding nurseries (47%) were allocated time each month or year to complete CPD. This is 10% lower than that reported by the control group at baseline despite control nurseries' CPD budgets remaining constant (average mean: baseline £1,673, endpoint £1,669). Time allocated to professional development varied across nurseries between set numbers of hours per week, a certain time of a specific day each week, to more liberal approaches of 'as and when needed'. INSET days were commonly used for training where necessary. One nursery had found that due to staffing changes, little CPD had taken place that year. Survey data suggested that limited finances were a barrier to future maths-related CPD. Additional training was often only possible if the budget allowed it, and in some nurseries additional training was undertaken in the staff's own time. Two nurseries

stated specifically that they could only access free or funded training—or they had to fundraise—and four stated that they had limited or no budget currently.

Drawing on the results of the primary outcome measure of this impact evaluation, the effect of the MC programme in the intervention group was over and above any maths-related CPD undertaken in control nurseries (which may or may not have had some impact on staff's maths-related classroom practice and children's maths attainment).

Summary of key findings

- Staff across all nurseries reported maths to be a part of everyday practice and was delivered as continuous provision.
- At baseline, more control nurseries reported having completed more maths- or numeracy-related and non-maths related CPD in the previous two years in comparison to intervention nurseries. Control nurseries also reported having a higher budget dedicated to staff CPD and training, and more control nurseries had plans to implement 'other' CPD during the trial in comparison to intervention nurseries.
- During the trial, some control nurseries reported completing maths/numeracy related and non-maths/numeracy related CPD or training, however among this group the number of nurseries that allocated staff time to complete CPD had reduced by 10% since baseline despite the average CPD budget remaining constant.

Fidelity and adherence

Data collected in relation to fidelity aims to describe the extent to which the MC programme was delivered as intended within intervention nurseries. This section also explores any issues with nurseries adhering to the MC programme as intended. The IPE research questions addressed within this section are:

RQ1 Is the MC programme delivered to MCs and DMCs with fidelity within both PVI and SN nurseries?

1.1: Are nominated staff (MCs, DMCs) accessing the available e-learning modules and the support as specified in the programme plan?

1.2: How effective and appropriate is the level of support and training (for example, content, coverage, dosage and duration) for MCs and DMCs?

1.3: What are the barriers for MCs and DMCs to engage with the e-learning modules?

1:4: What are the necessary conditions (facilitators) for MCs and DMCs to engage with the e-learning module and the one to one support?

Data to answer these research questions was provided from the e-learning logs and the interviews and focus groups with MCs, DMCs, managers, and practitioners.

Accessibility of the training

As part of the compliance analysis, the NDNA assessed the fidelity to the intervention within all nurseries receiving the MC programme. These data included nursery-level completion of the online courses (induction webinar to the programme and the subsequent two 'compulsory' online training courses). These data indicated that the vast majority of nurseries (56/66; 84.8%) were very engaged with and accessed the required induction and the online courses. Table 20 summarises nursery engagement with these components of the MC programme.

Table 20: Nursery engagement with the online induction and courses

MC and DMC complete induction	Very engaged	56 (84.8%)
	Partially engaged	9 (13.6%)
	Not engaged	1 (1.5%)
Completion by the MC of 2 courses:	Very engaged	56 (84.8%)

(1) developing mathematical confidence in the early years: the big ideas of number sense; and (2) developing mathematical thinking in the early years: shape space, measures and pattern—including characteristics of effective learning and sustained, shared thinking.	Partially engaged	7 (10.6%)
	Not engaged	3 (4.5%)

Overall, interview and focus group responses from MCs and DMCs within all 14 nurseries indicated that the MC programme induction webinar and online courses were accessible and useful. Feedback on the online courses was very positive, for example, all 14 nurseries stated that the delivery team was approachable, the quality of the training was high, and the level of support that was provided was 'absolutely appropriate' (DMC, nursery C):

> *'The materials and resources were really, really good; the activities were set out with links into the different [maths] concepts for children' (MC nursery C).*

No systematic differences between PVI nurseries and SNs were identified regarding MC programme delivery from the interviews and focus groups with MCs, DMCs, managers, and practitioners based on data from all 14 intervention nurseries.

The NDNA provided a phased entry to the training at the beginning and offered more induction sessions than it had originally planned, which they [NDNA] perceived worked well. It also ensured there were numerous different sessions that the MCs and DMCs could access, such as twilight sessions, as well as sessions through the day. The recordings enabled MCs and DMCs to look back on the webinars if they had missed something (a very welcomed feature by programme participants). The recordings were also intended to help in cases where a MC left a nursery and was replaced so the new MC was able to watch the induction and catch up on the training.

Effectiveness and appropriateness of support and training

MCs and DMCs from 12 nurseries (including both PVI nurseries and SNs) commented that the e-learning training content was all relevant and in line with the latest EYFS framework. They reported that the training resources and materials were 'exciting' (MC nursery C), 'self-explanatory' (DMC nursery G), 'informative' (MC nursery C), and 'straightforward' (DMC nursery N). Nursery C reported that the resources were 'fabulous', providing activity ideas to be shared in the nursery, that it had bought new resources and sorted old resources in light of the programme, and parents had provided more resources for this nursery. Moreover, MCs and DMS from *12 nurseries* reported that NDNA staff had provided regular contact and been available whenever the nursery needed to speak with them.

> *'I think their resources … were a really good starting point … self-explanatory but really focused on … the teaching bit and how to do that and then learn [what] to get from it … resources and worksheets [were] really straightforward and easy to follow' (DMC nursery G).*

In terms of which aspects of the content were particularly helpful or useful, MCs and DMCs from *seven* nurseries all commented that the terminology and glossary of terms introduced in the training were particularly useful.

> *'The [MC] programme is very useful in explaining [children's] maths learning, rather than assuming children should know this and that [maths knowledge]' (MC nursery M).*

One MC reported that she particularly liked:

> *'the terminology of things that we would potentially [be] doing, but we never knew what it was called and what the impact [was] on children' (MC nursery D.)*

The NDNA also believed that the most useful part of the online courses was the learning provided to increase knowledge and understanding of the different terminology.

MCs and DMCs from 11 nurseries (both PVI and SN) commented on the 'very good' mixture of resources and materials including readings, webinars and videos, audits, and core activities. The training facilitated 'insight' into children's mathematical capabilities (DMC nursery C). The MC in nursery C stated that the most valuable aspect of the programme was the core activities linked to specific mathematical concepts, for example, the importance of the composition of numbers, whereas previously their view on maths had been 'too narrow'.

> *'Videos were accessible, [they were] really relevant … so there's a bit of reading, a bit of video and a bit of the theory; it really mixed it all together for us' (DMC nursery G).*

The opportunity to attend the webinars flexibly, either the 'live' versions or the recorded versions was particularly appreciated by MCs and DMCs from *four nurseries*:

> *'We could make a lot of the webinars, I think the last one we missed, but I like that we could go back on and watch them as well, it just means I could watch that at my own time as well, so I like that' (MC nursery K).*

MCs and DMCs were asked to comment on the effectiveness and appropriateness of the training in terms of dosage, coverage, and duration. MCs and DMCs from nine nurseries commented that the training was at about the right level, and that dosage, coverage and duration were appropriate.

MCs and DMCs from these ***nine nurseries*** indicated that although a certain level of commitment was needed for the programme, the workload related to programme induction and training was manageable as long as they were prepared to commit adequate time for it. In terms of training dosage, however, MCs and DMCs from three nurseries indicated that they had missed training sessions due to COVID-19 breaks.

For four nurseries (J, G, D, and A) 'time' was a challenge for MCs and DMCs when they were in the dual role of leader and programme champion—where, for example, the nursery manager, headteacher, or room lead was also the MC (for example nurseries G, D and A).

> *'I'm being honest—I still haven't looked through all the material, the resources and I was hoping to do that in the summer holidays and as I'm looking through, I'm printing things out' (manager/MC nursery G).*

The same nursery expanded on the situation it was facing:

> *'It was tricky as well, because … she [MC] would have some time but we've both done it in our own time as well; I think that was one of the bits that's been hard' (DMC nursery G).*

However, these four nurseries had foreseen the possible challenges of the dual MC and manager role: the manager at one nursery, for example, was also the DMC and indicated that she had not nominated herself as MC due to potential lack of time to fulfil the role:

> *'I put myself as … definitely … backup; if I nominated myself [as MC], I would not get this all done' (DMC Nursery K).*

Barriers to engaging with training

The dominant view from ten nurseries of the barriers to engaging with the training were that there were no challenges whatsoever. However, as above, for four nurseries (J, G, I, and D) it became obvious that time management was a potential barrier for MCs and DMCs in the dual leadership and MC role. Working part time appeared to be another barrier to engagement with online courses. For example, the MC and DMC in nursery A both worked part time and reported this had an impact on their ability to attend live training sessions:

> *'I watched … the training videos because we just didn't have time after school [to attend the live training session], especially only being in two days … on days off time [taken] to watch the videos [to] try and catch up, I think that was quite a lot of extra work' (DMC nursery A).*

In terms of barriers to accessing e-learning resources, four nurseries indicated some minor technical issues. For example, the MC at nursery G admitted it took time to navigate the website. The MC and DMC in nursery I also encountered difficulties in accessing the platform.

The COVID-19 pandemic seemed to be a barrier for programme delivery for three nurseries, where the MCs and DMCs indicated that they had missed training sessions due to COVID-19 breaks.

Facilitators to engagement with training

As part of the compliance analysis, the NDNA assessed the fidelity to the intervention within all nurseries receiving the MC programme. These data included each nursery's engagement with the one to one support provided by the NDNA as part of the MC programme. These data indicated that the majority of nurseries 48/66 (72.7%) were 'very engaged' with the one to one support provided by the NDNA. Table 21 summarises nursery engagement with this component of the MC programme.

Table 21: Nursery engagement with NDNA's one to one support

Engagement with one to one support provided by the NDNA	Very engaged	48 (72.7%)
	Partially engaged	17 (25.8%)
	Not engaged	1 (1.5%)

MCs and DMCs from 11 nurseries indicated that an approachable delivery team was a condition for successful programme delivery, and they reported that the delivery team from the NDNA had been very encouraging, motivating, and not judgemental and that the support had been tailored to fit the various needs from different nurseries, which was appreciated by both MCs and DMCs in these nurseries.

MCs and DMCs from four nurseries commented on other necessary conditions: time and resource commitment from the start of the programme (nurseries E and H), preparation for the programme and awareness of the time and workload expectations (nurseries G and E), good time management skills such as doing small bits of training in a busy schedule (nurseries K, E, and H).

There were three main motivations behind nurseries signing up for the MC programme: maths was a learning area to improve on or explore further (nurseries F, NT, B and G), strong maths practice was already in place and nursery stakeholders wanted to get extra maths learning for staff as well as children (nurseries L, H, K and E), and extra PD activities in maths would benefit staff and children (nurseries D, I, M and J).

The NDNA provided an insight into how the training had changed for the programme delivery in this evaluation: in particular, the induction and online courses unpacked child development in greater detail than in the previous 'light touch' training.

Applicability of findings to the logic model

Training inputs in the logic model included the MC and DMC participating in relevant training (online induction and online courses and webinars) that aimed at equipping practitioners with a 'comprehensive understanding of the main areas of early years mathematical learning'. Programme resources in the logic model included the 700 activities within the resource bank of which ten were 'core resources', the monthly NDNA support (as a minimum), and the optional monthly webinar. The MCs' and DMCs' access to, participation in, and perceptions of the training and resource packages (induction webinar, online courses, and resource bank) was described extremely positively by all 14 nurseries. The training focused on children's mathematical development and maths terminology and language. MC and DMC attendance at the training and the consequent cascading of the content of the courses to other practitioners were features of the logic model, designed to improve early years practitioners' knowledge and understanding of children's mathematical development. In this sample of 14 nurseries, the unanimous view was that the training and resources were effective in achieving this aim. Therefore, the results from RQ1 fit with the training inputs and programme resources components of the logic model.

Summary of key findings

- The majority of nurseries (84.8%) were very engaged with the induction and the two online courses and with the monthly one to one support provided by the NDNA (72.7%).
- The dominant view from all 14 nurseries was that the MC programme training and support were delivered with fidelity within both PVI and SN nurseries.
- The dominant views from all 14 nurseries of the induction webinar and online training courses were extremely positive:
    - online courses were exciting, accessible, and useful;
    - they were of appropriate content, dosage, coverage, and duration;
    - training in the use of mathematical terminology was particularly useful; and
    - MCs and DMCs particularly valued the fact that they could catch up and rewatch the webinars at their own pace and in their own time.
- The dominant view from ten nurseries in terms of barriers to engagement with the training were that there were no challenges whatsoever.

- Views from up to four nurseries identified three barriers experienced in these specific nurseries:
    - time management was a challenge where the MC was also in a leadership role;
    - one nursery identified technical issues with the software; and
    - three nurseries identified the COVID-19 pandemic as impacting on the access to training.
- The dominant views from 11 nurseries on the necessary conditions for successful delivery of the training and support indicated that an approachable delivery team was the key condition:
    - the NDNA had been very encouraging and motivating but not judgemental.
- A range of views from up to three nurseries about other necessary conditions for successful delivery of the training included:
    - time and resource commitment from the start of the programme;
    - preparation for the programme and awareness of the time and workload expectations; and
    - good time management skills.

Implementation

Data collected in relation to implementation aims to describe the extent to which the MC programme was implemented by nurseries as intended by the NDNA. The IPE research questions addressed within this section are:

RQ2 To what extent is the MC programme implemented as planned within nurseries?

2.1. Do MCs and DMCs adhere to their roles as specified in the programme?

2.2. Do nursery practitioners implement the agreed action plans in their daily practice?

2.3. What are the barriers for MCs, DMCs, and practitioners to implement the programme in their classroom practice?

2.4. What are the necessary conditions for nursery practitioners to implement MC into practice?

RQ7 What is the perceived impact of the COVID-19 pandemic on the delivery of the MC programme?

Data to answer these questions was provided from the e-learning logs and the interviews and focus groups with MCs, DMCs, managers, and practitioners.

Adherence to roles

MCs and DMCs from all 14 nurseries commented that it was helpful to have a DMC, although it appears that nursery interpretation of the role of MC and DMC varied, and MCs and DMCs roles functioned slightly differently within nurseries in this sample.

We probed MCs', DMCs', as well as practitioners' understanding and interpretation of the MC and DMC roles and the functional mechanism within the nurseries; we identified three 'models' of the MC-DMC relationship, with Model 1 and Model 2 being the dominant types of relationship within six nurseries each.

Model 1 (six nurseries)

In Model 1 (***six nurseries***), MC and DMC partnership roles had no significant role and responsibility distinctions between them (nurseries K, E, H, B, J, and C). MCs and DMCs in these nurseries (both PVI and SN) attended the training sessions and led the MC programme implementation in different places or classrooms. In this type of relationship, MCs and DMCs supported each other, shared ideas and thoughts about children's mathematical development (DMC nursery C), discussed children's assessment, changed the environment to make it more maths orientated (DMC nursery C), acted as mentors for other practitioners in the nursery, and implemented the programme collaboratively.

> *'We can both work together and then, if I'm in another [nursery] room, I know that she's [DMC] going to do it, or if she's [DMC] off, I can go and do things. So it's nice to have that support [from each other]' (MC nursery H).*

> *'I won't ever do any of that stuff without telling [the DMC] what I'm doing so, realistically, even though I'm the champion and [they] are the deputy, there's not that much distinguishing between us really. There is always thorough discussion' (MC nursery B).*

Model 2 (six nurseries)

In Model 2, the MC and DMC shared roles and responsibilities (nurseries G, N, D, F, L and M; both PVI and SN). They both attended training sessions but they covered different tasks, for example, one was responsible for designing the action plan and overseeing the paperwork and one was responsible for making sure the action plan was implemented and embedded in the nursery rooms (for example, nurseries N and F). In this type of relationship, both MC and DMC functioned in a leadership role, and both were responsible for the programme but covered different elements. In one case, the DMC was the 'bridge' between the MC and the practitioners 'on the ground' and helped with making the programme 'more accessible'.

> *'I'm [MC] working on different things and then you're [DMC] making sure that that's getting transferred' (MC nursery N).*

Model 3 (two nurseries)

In Model 3, the MC and DMC had hierarchical roles: the MC leading but with MC and DMC working together to share the workload of the MC role (nursery A and I; both SN). The DMC functioned in a supportive role while being the 'leg man', the one to whom the MC was delegating MC role-related tasks (nursery I). In this type of relationship, the MC and DMC may or may not have attended the full training but the DMC was not involved in any of the planning stages.

Other observations

The NDNA observed differences in programme set-up between PVI and SN nurseries due to SNs having smaller teams whereas PVIs had a lot more staff—but also a lot more absence due to the pandemic. In terms of the DMC role, the NDNA perceived that everybody nominated by the nursery took on this role. In SNs, they looked at this more strategically: who was best to nominate? In some cases, this was the headteacher—a reliable presence with the knowledge to support implementation of the programme. In PVIs, both the MC and the DMC tended to be a practitioner. Also, it was apparent that PVIs had a higher staff turnover and in instances where the MC left, the DMC took on that role. In Robinson-Smith et al. (2018), the NDNA noted that staff turnover impacted on the delivery of the programme when the MC left the setting and there was no one else who could fulfil the role. As part of the pilot study (Appendix E), the NDNA trialled the role of the DMC and found it to be a successful additional support. The purpose of the DMC was to ensure the continuation of the MC programme within a nursery should the MC leave.

In terms of the CACE analysis using the red, amber, and green ('RAG') rating scores, there was a slight difference between type of nursery. SNs tended to be a little lower in the number of steps in the RAG they managed to complete. The NDNA was surprised about this as it had thought the structure of schools would enable them to participate more, for example, in INSET days, which meant they were out of the nursery room to enable them to do things.

The NDNA also observed that there had been a lot more engagement with the programme within the evaluation context than in the programme delivery for the previous evaluation (Robinson-Smith et al., 2018). It concluded that this had been due to a change to a focus on supporting children's mathematical development and changing practitioners' understanding of children's maths development. In the last evaluation, the focus had been on practitioners' understanding, but the new focus had led to greater engagement. The NDNA perceived that overall confidence and practitioners' own mathematical skills had developed. All but two nurseries (from entire cohort of 66 intervention nurseries, that is, 64 nurseries) had said this to the NDNA, but they said staff had scored themselves highly at the beginning. The NDNA perceived the DMC role to be a major role in the programme, which had been extremely positively received. Previously, if the MC dropped out there was a gap until a new MC was appointed and this had led to very good programme continuation.

Implementing change

Interview and focus group responses indicated differences between nurseries in terms of action plan implementation. Four followed the action plan closely (nurseries B, E, C, and M).

> *'We've used everything from the spinners to the coins; everything's been used in the nursery and also in the directed activities' (MC).*
>
> *We started with our staff meeting; we talked through the programme … then we printed off all the core activities, so we did all ten—we put it into a folder along with all the resources that were available … and then we planned in each week on our plan and as we went through those ten [activities] we revisited some of them that we hadn't quite got right. I went back and did it again and so that's how it's being embedded' (MC nursery B).*
>
> *'At the moment we're doing the core activities and quite a lot of them are done at the morning group time, as well as … the shop being implemented … and that's working really well' (MC nursery E).*

Eight nurseries reported that they implemented the action plans with various levels of adaptation and flexibility (nurseries L, H, D, N, A, J, C, and F).

> *'Some things [in the action plan] we were still working towards a little bit or we just wanted to change it back to extend it a bit more so the actual plan was pretty good, we needed some adaptations' (MC nursery L).*

The MC from nursery H stated that in this nursery the action plan was embedded in practice but that they used professional judgement to adapt things. The MC in nursery C stated that although they had followed the action plan, and adapted it, they felt that they had missed an opportunity to revisit it during the implementation.

Two nurseries (G and H) implemented part of the action plan over a shorter period of time.

> *'In regard to the core activities, we maybe [implemented them over] two months but we've totally embedded [it, as recommended by] the training' (MC, nursery G).*

Three nurseries (N, G, and A) encountered staffing issues such as staff shortages or absence due to sickness, which either delayed their action plan implementation by months or disrupted the implementation in the middle.

> *'The start of this year kind of hit with quite a lot sickness, which put us at a bit of a disadvantage with it [the implementation] (MC nursery N).*
>
> *'This [COVID] year has been so hard … I know we're quite behind still on what we need to do, but it's just with staff off, children off, it's just really hard' (MC nursery A).*

Only one nursery reported that it had not implemented an action plan at all due to difficult times. This was because the nursery was going to be removed from the school and they had only engaged mainly for programme evaluation and assessment tasks.

> *'We've not been able to deliver it [the action plan] because we are actually axing the nursery' (MC from nursery I).*

The NDNA perceived the plans to be very helpful for the nurseries in terms of assessing potential areas for improvement.

Barriers and facilitators to implementation

We explored the barriers encountered by nurseries in incorporating the MC programme into room practice. Interview and focus group responses from MCs, DMCs, and practitioners highlighted the main challenges: three nurseries reported staffing absence and staff shortage (N, G, and A); two nurseries reported workload issues (A and N), and three nurseries reported COVID-19 had impacted the MC programme implementation (G, I, and N).

> *'I would say no difficulties … but time has been a factor, [although] we have been quite systematic and quite organized, but obviously we have [had] staff absence because of COVID. It sometimes has [an impact] on the schedule of what we were planning' (MC nursery G).*

In terms of the necessary conditions for successful programme implementation into room practice, we identified the following conditions from MCs, DMCs, and practitioners' interview and focus group responses from seven nurseries (E, H, L, N, F, B, and I):

- staff awareness of the MC programme—every staff member being on board with the implementation (for example, nurseries H, N, and B):

'I think that's one of the most important things—to make sure staff know exactly what's going on and [be] on board, and it's not just run through them, because we've got to put a lot of time into it' (MC nursery N);

- MCs and DMCs are committed to the programme and good at time management and action planning—MCs and DMCs demonstrating and modelling the programme; they are part of the implementation team; they use the resources and act on core activities quickly (for example, nurseries F, B, and H):

  'You have to be committed definitely to start with, you have to go into it with your eyes open and make sure you understand the amount of work that's involved … make sure you're organised, and you've got like a maths folder' (MC and practitioner nursery B); and

- continuity of programme implementation in practice (for example, nurseries E, G, and L):

  'You've got to continue it, you can't just do it for the programme, and then finish … I got all the ten fixed activities, but I've also got all the other activities that we can just keep using over and over again' (MC nursery L).

Applicability of findings to the logic model

The second inputs component of the logic model included the range of programme activities in terms of the practitioners' implementation of the elements (for example, a nursery-specific action plan that was continuously reviewed by the NDNA) and using MC resources and tools, (for example, trackers to observe and monitor child progress and online resources aimed at improving maths provision). Most of the outputs in the logic model were perceived by the stakeholders interviewed to have been achieved. The MC programme implementation in the evaluation reached both the PVI and SN nurseries. In the IPE sample of 14 nurseries, there was a 50% split in type of nursery and no obvious differences in implementation were noted. In all nurseries in the sample, the MC disseminated and ran the programme within their nursery. In this sample of 14 nurseries, the unanimous view was that the role of the DMC was helpful, although there were three 'models' of MC-DMC relationship observed, two of which appeared more dominant and more widely adopted. In all models, the DMC worked with the MC in implementing change. All practitioners within a nursery implemented the action plans in daily practice with varying degrees of flexibility. Thirteen nurseries stated that they had implemented the action plan effectively with various degrees of flexibility and adaptation. Very few barriers to implementing the MC programme were identified and by only three nurseries. The dominant view (11 nurseries) was that there were no barriers to implementing the programme. Three necessary conditions were identified by seven nurseries. The implementation of the programme elements was perceived to have been successful in the majority of nurseries (11) and, therefore, these results fit with the programme activities (practitioner implementation) input component of the logic model.

Summary of key findings

- Three 'models' of the MC-DMC relationship were identified with Models 1 and 2 being the dominant types of relationships in this sample of nurseries. The three models were:
  1. a partnership (six nurseries)—no significant role and responsibility distinctions between MC and DMC;
  2. split roles (six nurseries)—responsibilities shared between MC and DMC; and
  3. hierarchal (two nurseries)—the DMC subordinate to the MC.
- The dominant view of the DMC role from all 14 nurseries was that it was helpful to have a DMC.
- A range of views on the implementation of the action plan was observed although only 13 nurseries implemented it:
  - four nurseries followed the action plan closely;
  - eight reported that they implemented the action plans with various levels of adaptation and flexibility;
  - three reported a delay in the action plan implementation;
  - only one nursery reported that it did not implement an action plan; and

  - the NDNA perceived the plans to be very helpful for the nurseries in terms of auditing highlighted areas for improvement.

- The dominant view was that there were no barriers to implementing the MC programme in nurseries (11 nurseries).
- Barriers to implementing the MC programme, in up to three nurseries, included:
  - staffing absence or staff shortage (three nurseries);
  - workload issues (two nurseries); and
  - the COVID-19 pandemic (three nurseries).
- Seven nurseries identified three necessary conditions for successful implementation of the MC programme:
  - staff awareness of the programme;
  - MCs and DMCs being committed to the programme and good at time management and action planning; and
  - continuity of programme implementation in practice.

Stakeholder perspectives

Data collected in relation to stakeholder perspectives aims to describe how nursery staff perceived the effectiveness of the MC programme in relation to the impact it had on their own ability to teach maths and also any observed impacts on the children as a result of their change in practice. The IPE research questions addressed within this section are:

RQ3 What are the different stakeholders' viewpoints on the MC programme?

The data sources for RQ3 were the interviews and focus groups with MCs, DMCs, managers, and practitioners.

Perceived impact on staff

Thirteen nurseries reported that there had been perceived improvement of confidence in delivering maths activities from the MC, the DMC, as well as practitioners.

> *'Everybody's confidence has soared … belief in myself [has increased] immensely' (DMC and practitioner nursery C).*

MCs and DMCs from four nurseries (H, L, E, and M) mentioned they were already very confident in their provision of maths.

> *'For me personally I've always been confident, so I don't think it's changed, but literally maths is my thing.'*
> *'I feel like it's [maths] more [at] the forefront of my mind' (MC and DMC nursery E).*

The MC or DMC from nine nurseries (F, D, I, N, A, J, G, C, and B) reported that, at a personal level, they had become more confident in delivering maths activities. The practitioner in nursery M stated that they had been initially 'daunted' by the thought of maths but due to the 'amazing' work of the MC and DMC in that nursery in sharing thoughts and ideas, practitioner knowledge had been widened to assist in engaging the children in the maths activities.

> *'[We are] most definitely much more confident in our own abilities, I think, and also in the abilities of the staff as well' (MC nursery B).*

> *MC: 'Definitely [I am] more confident … I am so much more confident in maths.' DMC: 'I think, in general, it [the MC programme] has just given us more ideas and made us more aware of things you can do … I think it's given us more confidence to do it' (MC and DMC nursery N).*

Nine nurseries mentioned that there had been observed improvement in terms of the confidence of practitioners (B, N, J, G, L, H, F, C, and K).

> *'They [practitioners] are more confident in teaching maths and also developed in maths within the continuous provision so spontaneously trying to incorporate maths within conversations' (MC nursery G).*

Two nurseries did not specifically mention improvement in confidence but observed more maths activities being undertaken by practitioners (D and M).

Perceived impact on children

Eleven nurseries reported delivering more maths activities, providing more resources, or covering more maths areas. Such changes to the learning environment are likely to impact on children's learning. At the practitioner level, use of mathematical language was increased (nursery K, B, A, and F). In addition, seven nurseries reported increased awareness and understanding of maths and children's learning (nurseries B, L, E, K, F, C, and A).

Three nurseries reported engaging children more with maths (B, C, and N); and one mentioned their 'eyes were opened' due to the resources (nursery D).

We explored the perceived impacts of the MC programme on child outcomes from MCs, DMCs, and practitioners. Six nurseries (F, L, H, I, C, and B) observed children engaging more with maths materials and doing more maths-related activities; the daily routines in the nurseries changed to use maths throughout the day and the new routines came 'naturally' to the children (DMC nursery C); the children used more mathematical language and visitors and parents reported improvements.

> *'They've [children] enjoyed those [maths] activities and … once we've done the activities with children, sometimes we've left the materials out and the children have taken them on themselves … it's been nice to step back in to watch the children and develop confidence in using some of the materials with their friends without an adult there as well' (MC nursery F).*

> *'They [the children] are using it [maths terminology] all the time, but even when [we're] out and about with them now [they're] noticing shapes and naming things and numbers' (MC nursery L).*

Five nurseries (L, N, B, K, and E) commented that children were using more mathematical language.

> *'I think it's had a brilliant impact on the children; they're using mathematical terminology a lot more; they're incorporating it into their play a lot more' (MC nursery L).*

One nursery commented that school visitors have also noticed big differences in children's maths skills.

> *'We've noticed a big difference in their mathematics and when we've had visitors in, they've commented on their maths or on their counting' (MC nursery F).*

Three nurseries (H, C, and B) commented on parents' feedback regarding children's improvement in maths.

> *'Parents responses [are] really positive; they said that they've noticed how well the children are doing … they are much more confident, they are using it [maths terminology] all the time' (MC nursery K).*

The Early Childhood Environment Rates Scales-III (ECERS-3) and the ECERS-E

ECERS is scored by the observer assigning a descriptive value on a scale from one to seven ('inadequate' to 'excellent') that describes the quality of the early childhood environment for each subscale. A final score can be tabulated as an average (mean) of the scores of the subscales and items that were used in the tool. The data from these scales collected in a small subsample of nurseries (n = 3) provides limited support for the improved quality of maths provision within these three nurseries—a snapshot of the maths learning environment in these three—however, the results are not applicable to all 66 intervention nurseries. Overall, all three nurseries' maths-related learning environment—as measured by the ECERS-3 item 23 ('maths materials and activities'), item 24 ('maths in daily events'), and item 25 ('understanding written numbers') and ECERS-E item 7 ('counting and the application of counting'), item 8 ('reading and writing simple numbers'), item 9a ('mathematical activities: shape and space'), or item 9b ('mathematical activities: sorting, matching and comparing')—improved between baseline audit and follow-up audit (average mean score: nursery E, 3/7 at baseline, 5/7 at follow-up; nursery O: 3/7 at baseline, 5/7 at follow-up; nursery P: 4/7 at baseline, 5/7 at follow-up,). These results triangulate with the stakeholder perceptions from the 14 IPE nurseries (of which one was an ECERS observation nursery).

Applicability of findings to the logic model

In the causal mechanism in the logic model, one of the outcomes is increased practitioner confidence in maths and in delivering maths activities with the nursery children. The dominant perception from 13 nurseries was that there had been

perceived improvement of confidence in delivering maths activities from the MC and the DMC as well as practitioners and, therefore, the results from RQ3 support the logic model in terms of practitioner perceptions.

Summary of key findings

The dominant views from 13 nurseries were that there had been perceived improvement of confidence in delivering maths activities—from the MC, the DMC, as well as practitioners.

- MCs and DMCs from nine nurseries reported that, at a personal level, they had become more confident in delivering maths activities.
- The majority view from 11 nurseries reported an improved maths learning environment:
    - undertaking more maths activities, providing more resources, and covering more maths areas;
    - increased use of mathematical language; and
    - increased awareness and understanding of maths and children's learning.
- Six nurseries observed children engaging more with maths materials and doing more maths-related activities:
    - the daily routines changed to using maths throughout the routine;
    - the new routines came 'naturally' to the children;
    - the children used more mathematical language; and
    - visitors and parents reported improvements.

Adherence to trial procedures

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

RQ4 To what extent does the MC programme impact evaluation process adhere to the plan?

4.1 Do nursery MCs and DMCs meet the specified recruitment criteria for the MC programme?

4.2. Does the children and family recruitment process adhere to the recruitment strategy?

4.3 Do baseline and outcome test administrators (teachers or independent research assistants) effectively and appropriately evaluate children's maths attainment?

4.4 Are there any sample attrition effects and how that might affect the estimates of the impact of the MC programme?

To address these research questions, we summarise data gathered within the impact evaluation, from baseline surveys, and from staff interviews and focus groups.

Staff and child recruitment

All 66 intervention nurseries were assessed by the NDNA for their fidelity to the intervention. As detailed in

Table 14 in the CACE results section above, all 66 nominated a MC and DMC qualified to at least level 3. This data confirms that all nominated MCs and DMCs met the specified recruitment criteria for the MC programme.

Nurseries were requested by the evaluation team to distribute information sheets and consent forms to the parents and carers of all eligible children and were discouraged from 'self-selecting' which parents or carers they approached to

participate, or from 'self-selecting' from the returned completed consent forms the ten children whom they preferred to complete the ASPECTS assessment. Parents and carers in most nurseries were given the option to complete the consent form either online or on paper (see Table 22 for method of consent completion by randomised group).

Table 22: Method of parent and carer consent form completion, by randomised group

Method of consent completion	Intervention		Control		All nurseries	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)	n/N (missing)	Count (%)
eConsent only eConsent and paper Paper only	66/66 (0)	38 (57.6) 12 (18.2) 16 (24.2)	68/68 (0)	26 (38.2) 22 (32.4) 20 (29.4)	134/134 (0)	64 (47.8) 34 (25.4) 36 (26.9)

Qualitative data gathered from the baseline surveys indicated that nurseries generally reported being able to adhere to the recruitment strategy. For some, recruitment was straightforward and staff did not have issues obtaining consent from parents or carers who were generally enthusiastic about the trial. This was a recognised theme during interviews.

> *'We have 21 children [who] were eligible and 20 of them came back [with consent forms]. So we [didn't] really chase that' (MC nursery N).*

> *'Lots of parents [are] really keen on the programme; parents talk to us quite often about it … they're very interested to know the outcomes as well, so it's really good' (MC, DMC nursery E).*

In the baseline survey, other nurseries shared experiences that suggest gaining parent or carer consent was harder or that they had to make additional efforts to aid child recruitment. There were concerns over the amount of information parents or carers had to read both in the information sheet and the consent form. It was discussed that making this more succinct would have aided recruitment as some either did not read it properly or chose not to participate due to this. In contrast, some parents and carers required additional information or had additional questions before agreeing to sign the consent form.

It was reported that some parents or carers were interested in participating but that this sometimes failed to translate into a signed consent form. It was very common for them to need reminders to complete the form. In some nurseries, one or two reminders were required, however, in others, it was not until staff members went through the forms with the parents or carers that consent was obtained. Doing this was also a strategy employed in nurseries with parents/carers who needed extra support completing the forms (either due to language barriers, reading difficulties, or technological barriers). A small number of nurseries reported that parents and carers had concerns over data sharing and information being kept for five years and chose not to give consent due to this.

Two nurseries noted that it was to their disadvantage that programme information was given to parents and carers either before the child had officially started at the nursery or outside of term time as parents and carers were not checking their emails as regularly and were less engaged with communication from the nursery.

Test administration

The baseline ASPECTS assessments were administered by a practitioner within each participating nursery prior to randomisation. Where nurseries lacked capacity to complete these, a member of the evaluation team visited to assess the children. This occurred in eight nurseries (52 participating children). At post-test, trained, independent Ras visited all 134 randomised nurseries and administered the post-test to 1,176 children (97% of those post-tested children). Ras were extensively trained by the evaluation team on how to administer the ASPECTS post-test to participating children prior to their first visit. A member of the evaluation team visited each RA on their first or second nursery visit to quality assure the administration of the post-test (nine Ras were joined on their first visit and two were joined on their second visit). No significant issues were identified. Some children were absent on the days or times when Ras were scheduled to visit the nurseries and conduct the post-tests. Where feasible, Ras made a second or third visit to try to assess all randomised children. Where it was not possible for an independent RA to visit a nursery again, practitioners at the nursery were requested to complete the ASPECTS post-test with the remaining children. The post-test was administered

by a nursery practitioner or teacher for 26 children (2.2% of the post-tested children: intervention, *n* = 16; control, *n* = 10) across 21 nurseries (intervention, *n* = 12; control, *n* = 9). Reasons for an RA not being able to complete the post-test, and the nursery being asked to complete the remaining assessments Instead, included:

- the nursery being remote or difficult to travel to and only having one or two children left to assess;
- the child not attending when the additional visit was scheduled to assess them at the end of the school term or assessment period;
- additional visits being booked but then cancelled due to either the child being on holiday or the assessor being unwell or testing positive for COVID-19;
- the nursery being unwilling to book in another visit due to not knowing when the child would be present;
- the nursery being unable to accommodate another visit due to a COVID-19 outbreak in the nursery; and
- the child not wanting to start or continue the assessment with the RA during the initial visit and the nursery agreeing to try to complete the assessment.

A total of seven children (0.6% of all post-tested children) were administered the post-test by a member of the evaluation team. This occurred as the RA scheduled to complete the assessments was delayed due to a national rail strike and a member of the evaluation team was available at the nursery as they were due to complete a quality assurance observation of the assessments.

Impact of attrition

Of the 134 nurseries that were randomised, four withdrew from delivering the MC programme but were retained within the evaluation. There were no formal child withdrawals from the trial. Of the 1,304 children who completed the baseline assessment, 1,209 completed the post-test. This equates to child-level attrition of 7.3%. Baseline characteristics for the 1,304 randomised children and the 1,209 children included in the primary analysis are presented in Table 10 and Table 11 for visual comparison. No notable differences were observed. Movement of children out of nurseries was very low: of the 1,304 children in the randomised sample, 63 children (4.8%) left the nursery during the trial; of the 694 children from randomised PVIs, 39 children left (5.6%) and of the 610 children from randomised SNs, 24 (3.9%) left. From our checks with nurseries before post-testing, we were informed that 55 children had left and during post-testing we were made aware of eight more. Due to this, and in consultation with the EEF, we decided not to approach any new nurseries and instead to focus resources on assessing as many of the participating children who remained in their original nurseries as possible.

Summary of key findings

- The delivery of the impact evaluation largely adhered to the trial protocol (Robinson-Smith et al., 2022).
- All MCs and DMCs met the relevant eligibility criteria (qualified to at least level 3) and nurseries successfully obtained parent or carer consent for a sufficient number of children to participate in the trial.
- Practitioners completed the baseline assessment with 96% (1,252 of 1,304) of participating children and 4% (52 of 1,304) were assessed by an assessor from the evaluation team; 97% (1,176 of 1,209) of the post-tests were administered by a trained independent RA, 2.2% (26 of 1,209) by the nursery practitioner, and 0.6% (7 of 1,209) by a member of the evaluation team. From this data, we can conclude that test administrators were able to effectively and appropriately evaluate children's maths attainment.
- The impact evaluation experienced minimal attrition. All 134 randomised nurseries were retained within the evaluation for post-testing, despite four intervention nurseries withdrawing from programme delivery. Child-level attrition was low, at 7.3% overall, as fewer children than anticipated (based on the first MC effectiveness trial) left their nursery between baseline and post-test (63 of 1,304, 4.8%). No notable differences were observed between the baseline characteristics for the children randomised and those included in the primary analysis. Therefore, we deem the risk of sample attrition effects to be low. The trial was therefore well-powered and we can have confidence in the validity and reliability of the impact evaluation results.

Impact of COVID-19 on programme delivery

Drawing upon the experiences of key nursery staff and the NDNA, this section describes the extent to which the COVID-19 pandemic impacted on the delivery of the MC programme within intervention nurseries. The IPE research question addressed in this section is:

RQ7 What is the perceived impact of the COVID-19 pandemic on the delivery of the MC programme?

The data sources were intervention MCs and managers or headteachers responses to the endpoint survey, the focus group with the NDNA, and the interviews and focus groups with the MCs, DMCs, managers, and practitioners at 14 nurseries.

Nursery staff

In the endpoint survey for Maths Champions, 15 of 41 respondents (36%) in the intervention group indicated that COVID-19 had had an impact on the implementation of the programme. Impacts noted were staff absence (13 nurseries), children's absence (four nurseries), children's lack of focus (one nursery), and lower parental engagement (one nursery).

In the endpoint survey for headteachers and managers, 15 of 41 respondents in the intervention group (36%) also indicated that COVID-19 had had an impact on implementation of the programme; four of these nurseries were those which withdrew from the intervention. The 11 nurseries that did not withdraw but stated COVID-19 *did* impact on the implementation gave the following reasons:

- an impact on the attendance of staff and children:
    - children did not initially attend regularly when nurseries re-opened (two nurseries) and staff had to focus on the children when they returned (one nursery); and
    - staff had to cover for colleagues (in other rooms) who were ill with COVID-19, and there were issues with staff consistency (six nurseries); and
- an impact on planning and paperwork:
    - planning and paperwork had to be re-scheduled and, in some cases, completed at home; and
    - time was more constrained for the implementation.

However, one nursery stated that, despite all areas of provision being affected by COVID-19, due to the hard work and commitment of the staff to the MC programme their usual implementation was maintained.

All four nurseries that withdrew from the programme stated that the sole reason was the impact of COVID-19 on staffing issues, including staff absences and understaffing.

The COVID-19 pandemic was perceived to have impacted on (a) training: as mentioned before, MCs and DMCs from three nurseries (K, G, and I) reported that they missed webinars because of COVID-19; and (b) implementation: MCs and DMCs from six nurseries mentioned that COVID-19 had impacted their programme implementation due to staff absence due to sickness, covering for sick colleagues, workload issues, and other COVID-19 issues such as the necessity for enhanced cleaning (nurseries D, N, G, I, A, and H).

> *'I think [the workload is manageable] if there wasn't COVID and the lack of staff, because at the moment, staff are really hard to find' (MC nursery N).*

However, only one nursery was unable to deliver the action plan (nursery I) due to various difficulties including COVID-19:

> *'Sadly we've had staff off with COVID and then their own children off [with COVID] so they've been off for long periods than we would normally anticipate' (MC nursery I).*

Only one nursery delayed the implementation due to COVID-19 (H):

> *'A six-week period … put us behind … for everything and it was a big catch-up' (MC nursery H).*

With regard to children, very few nurseries reported impacts as a consequence of the COVID-19 pandemic: one nursery reported children had issues around attachment due to separation from parents (nursery B), two nurseries reported issues with children's attendance (nurseries I and A), one nursery reported issues with social skills (B), one nursery

reported that children took time to get used to routines and group activities following their return to nurseries (L), one nursery reported the necessity to support children and help them catch up (nursery F), and one nursery reported children's absence as being an issue for the children but also reported that the MC programme helped them catch up (N).

The National Day Nurseries Association (NDNA)

Overall, the NDNA did not perceive the impact of the pandemic to have been as bad as it had predicted and was surprised by the high retention rates of nurseries and the low attrition rates of children. However, the NDNA identified three areas—relating to training, absence, and children—where it felt the pandemic had had a measure of impact on the programme.

Training

Delivery of the training was always online so there were no major changes to this aspect of the programme except to add disclaimers in terms of government guidelines about the pandemic and to be more flexible about completion dates due to lockdowns and staffing issues. The NDNA perceived the online delivery to be a strength of the programme for two reasons: first, MCs and DMCs did not have to be released to go to a venue to receive the training; second, because the training was online the NDNA was able to work through the programme with the MCs and DMCs when required. Induction attendance was high, with MCs and DMCs having attended from 85% of nurseries.

Absence

Staff absence due to sickness was perceived by the NDNA to be a 'massive barrier' to the week-on-week continuum within the programme and felt that this had led to the disappointment of some MCs who thought they could have done more if they had had more time. Absences in January 2022 were particularly high due to the impact of COVID-19.

Children

A few nurseries noted that they perceived children had been affected by the pandemic. The NDNA was concerned that there would be developmental impacts, however, NDNA perceptions were that children's learning and understanding (language skills and problem-solving skills)—attributed to the programme—improved and the understanding of practitioners increased. Nurseries said it had been good to focus on something other than the pandemic.

Summary of key findings

- Overall, COVID-19 did not appear to have a significant impact on programme implementation in most nurseries.
- A small proportion of nurseries reported in the surveys and in the interviews and focus groups that there was an impact due to the pandemic on staff and child absences. Although the NDNA felt that this had been a barrier to consistent implementation, this perspective was not borne out by the views of most nurseries.

Cost evaluation[8]

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

We estimated the average cost per child per year for nurseries implementing the MC programme following the EEF costing guidance issued in 2019 (EEF, 2019). A year is defined as a year of implementation. This may not align with the calendar or academic year. This costing model estimated costs based on the mean number of eligible children per nursery included in the evaluation (n = 34). Given that this is a staff CPD programme, all children attending a nursery are likely to be impacted, thus, the per child costs would be dependent on the size of the nursery and could be reduced if they are spread across a greater number of children.

Table 23 details the resources needed to implement the programme as per the ingredients method (Levin et al., 2017). The main cost for implementing the programme is the cost of training. Other costs identified include staff cover and additional materials, however, these are categorised as optional rather than mandatory as only a proportion of nurseries utilised these.

Table 23: List of resources—'ingredients'

Category		Item
Personnel for preparation and delivery		Nursery staff—Maths Champion
		Nursery staff—Deputy Maths Champion
Personnel for training	Trainers	National Day Nurseries Association—early years expert
		National Day Nurseries Association—project lead
		National Day Nurseries Association—project coordinator
	Trainees	Nursery staff – Maths Champion
		Nursery staff – Deputy Maths Champion
	Project management and admin	National Day Nurseries Association – project coordinator and administrative staff
Facilities, equipment, and materials (prerequisites)		ICT equipment, e.g., laptop, computer, or Ipad and internet connection
		Learning resources to support mathematical learning
Optional Extras*		Additional learning resources to support mathematical learning (as needed)
		Staff cover (as needed)

*Some nurseries may need to purchase additional resources if they do not already have them in place as part of their usual provisions. Some nurseries may need to provide staff cover to release MC's and DMC's for training activities to ensure adequate ratios are maintained.

Prerequisites

In order take part in the MC CPD programme, nurseries needed to have access to ICT equipment such as a laptop or PC and an internet connection. To our knowledge, no nursery purchased such equipment specifically for the purpose of taking part in the evaluation and we anticipate that most nurseries would already have this equipment should the programme be rolled out.

While implementing the MC programme, nurseries were expected to utilise their existing learning resources. A survey was conducted at the outset of the evaluation to understand what existing resources nurseries were making use of. Nurseries reported the use of ICT facilities (computers, internet, printer laminator), stationery, and existing learning resources including shapes, blocks, counting objects, cards, and books.

[8] The previous version of this report, published in July 2023, contained inaccurate cost figures. This updated version rectifies those inaccuracies and includes a revised analysis.

Time

The MC programme is intended to be delivered as a whole-nursery, all-staff approach and, as such, is delivered to some extent by all staff throughout the day. However, each nursery nominates a MC and DMC and these individuals are required to spend some time in training or CPD. Given this, the time costs are largely front loaded and associated with start-up training and continued development activities undertaken during the course of the implementation period (seven months in this evaluation). Training was provided to both the nominated MC and DMC per nursery via a combination of webinars, online courses, and remote support sessions, some of which were optional.

Surveys were distributed to MCs and DMCs following their start-up training and again at the end of the implementation period. They were asked to estimate the amount of time they had spent on the start-up and ongoing activities, both within working hours and in their own time, and this data is summarised in Table 24 and Table 25.

During the initial set-up of the programme, MCs spent, on average, 11.19 hours engaging in start-up training activities during working hours; the corresponding figure for DMCs was, on average, 6.55 hours. Around two thirds (68%) of MCs and one third (36%) of DMCs reported undertaking work during their own time. This averaged 6.27 unpaid hours for MCs and 7.76 for DMCs.

Following the initial start-up phase, MCs spent, on average, 15.05 hours and DMCs 14.60 hours engaging in ongoing programme activities during working hours over the seven-month period. Similar proportions of MCs and DMs continued to undertake work in their own time, 65% and 33% respectively, with MCs reporting having worked an average of 5.41 unpaid hours and DMCs 1.33.

In addition to collecting data around staff time, we also asked nurseries whether it was necessary to arrange cover for the MC and DMC. Less than one third of nurseries reported staff cover utilisation at start-up: eight of 25 nurseries (32%) reported arranging MC cover and six (24%) DMC cover. This fell considerably at the follow-up survey with only 2 of 17 nurseries (12%) reporting the need for MC cover and only one (6%) reporting the need to cover the DMC. Given the timing of this evaluation, nurseries were asked whether cover was necessary due to COVID-19 and associated staff shortages: in half of the cases requiring cover this was the case. Given the variable uptake of staff cover, this is considered an optional cost that would be utilised on a nursery-by-nursery basis.

Table 24: Total time devoted by personnel for training, staff cover, and ongoing activities during working hours

		Year 1	
		Number of teachers	**Mean number of hours (SD)**
Start-up training	Maths Champion	1 per nursery	11.19 (5.87)
	Deputy Maths Champion	1 per nursery	6.55 (2.95)
Ongoing activities	Maths Champion	1 per nursery	15.05 (20.88)
	Deputy Maths Champion	1 per nursery	14.60 (25.28)
Teacher cover*	Maths Champion	1 per nursery	2.76 (8.20)
	Deputy Maths Champion	1 per nursery	2.33 (4.75)

* Hours of teacher cover was collected at the end of implementation so does include hours of cover for start-up activities.

Table 25: Total time devoted by personnel for training and ongoing activities outside working hours

		Year 1	
		Number of teachers	**Mean number of hours (SD)**
Start-up training	Maths Champion	1 per nursery	6.27 (6.36)
	Deputy Maths Champion	1 per nursery	7.76 (3.35)
Ongoing activities	Maths Champion	1 per nursery	5.41 (7.78)
	Deputy Maths Champion	1 per nursery	1.33 (2.42)

Financial costs

Training

All training took place remotely and was either in the form of a webinar, a one to one session, or through directed independent learning. A similar model could be used to roll out the programme.

Training of nursery staff was delivered by colleagues from the NDNA and involved an early years expert, project lead, and project co-ordinator. Nurseries were trained in groups. To deliver training to all nurseries allocated to the programme arm of the trial, the NDNA ran a total of 11 one-hour induction sessions. It also hosted five 30-minute webinars. The NDNA also provided monthly one to one telephone or video link support with nurseries receiving, on average, one ten-minute, one to one session per month for seven months (70 minutes per nursery). Nurseries also accessed IT and technical support from the NDNA and the NDNA has provided us with an estimate of the time and financial cost of this. Table 26 presents the breakdown of programme delivery costs from the perspective of the NDNA, these include intervention development, setting recruitment, and intervention delivery.

Staff cover

As detailed above, cover was used variably by nurseries. On average, across all nurseries providing data, £25.88 was spent on cover during the start-up phase and £78.37 during the implementation phase (

Cost ingredient	Start-up or recurring?	Nominal values		
		£ Year 1	£ Year 2	
Intervention Development	Start-up	£ 7,976.29	£0	
Setting Recruitment	Start-up	£ 8,965.66	£0	
Intervention Delivery (online training, 1:1 support, management and administration of the programme)	Start-up	£ 151,106.60	£0	
Total cost of programme delivery for all nurseries		**£ 168,048.54**	**£0**	
Total cost of programme delivery per nursery (66 nurseries)				
Total cost per pupil-school-year (34 children per nursery over 3 years)				

Table 27). Given the low number of nurseries reporting the need to pay for cover, this is considered an optional cost, which will vary on a nursery-by-nursery basis. Nurseries considering implementing the MC programme should factor this into their considerations.

Materials

Additional materials

In addition to being asked about their utilisation of existing materials, nurseries were also asked whether they were required to purchase any additional materials; where this was the case, they were asked to provide monetary costs. On average, nurseries reported spending £48.36 on additional resources during the start-up phase of the programme, ranging from £0 to £250, and an additional £66 by the end of the evaluation, ranging from £0 to £625. Additional resources purchased including items such as measuring jugs, sorting items, puzzles, rulers, tape measures, and stationery. Not all nurseries reported purchasing additional materials: only eight of 25 nurseries (32%) at start-up and ten of 17 (59%) post-implementation reported doing so, hence additional learning resources are considered an optional rather than mandatory cost for implementation. Costs associated with materials could represent both a start-up or

recurring cost depending on the frequency that resources need to be replaced. Costs are likely to be lower at start-up for nurseries that already have a wealth of relevant learning resources and higher for those that do not.

Overall costs

In total, NDNA estimates that the total programme delivery costs for the 66 intervention nurseries amounted to £168,048.54, which equates to £2,546.19 per nursery (total spent divided by the number of nurseries trained). This constitutes the largest portion of the overall cost. Additional costs were all optional and uptake varied depending on nursery. Such costs included additional materials and staff cover.

Table 28 presents the total cost per child per year over three years based on the trial costs. As stated above, this calculation assumes an average of 34 children per nursery and that no further top-up training is required after the first year. Additional costs could be incurred if nurseries need to replace or purchase additional learning resources. This cost would be at the discretion of an individual nursery but we have used the costs gathered through the nurseries included in this evaluation to provide an indicative estimate.

Table 26: Cost of the implementation of the programme, per mandatory ingredient

Cost ingredient	Start-up or recurring?	Nominal values			
		£ Year 1	£ Year 2	£ Year 3	Total
Intervention Development	Start-up	£ 7,976.29	£0	£0	£ 7,976.29
Setting Recruitment	Start-up	£ 8,965.66	£0	£0	£ 8,965.66
Intervention Delivery (online training, 1:1 support, management and administration of the programme)	Start-up	£ 151,106.60	£0	£0	£ 121,362.72
Total cost of programme delivery for all nurseries		**£ 168,048.54**	**£0**	**£0**	**£ 168,048.54**
Total cost of programme delivery per nursery (66 nurseries)					**£ 2,546.19**
Total cost per pupil-school-year (34 children per nursery over 3 years)					**£ 24.96**

Table 27: Cost of the implementation of the programme, per optional ingredient

Category	Cost Ingredient	Start-up or Recurring?	Nominal Values			
			£ Year 1	£ Year 2	£ Year 3	Total
Staff Cover (start-up) (optional)	MC cover	Start-up	£11.30	£0	£0	£11.30
	DMC cover	Start-up	£14.58	£0	£0	£14.58
Staff Cover (ongoing) (optional)	MC cover	Start-up	£47.64	£0	£0	£47.64
	DMC cover	Start-up	£30.73	£0	£0	£30.73
Materials (start-up) (optional)	Learning resources for children	Start-up	£48.36	£0	£0	£48.36
Materials (ongoing) (optional)	Learning resources for children	Recurring	£66.00	£66.00	£66.00	£198
Total cost of optional items per nursery			**£218.61**	**£66.00**	**£66.00**	**£350.61**
Total cost per child (34 children per nursery)						**£10.31**
Total cost per pupil-school-year (34 children per nursery over 3 years)						**£3.44**

Table 28: Total combined costs of training and optional extras

Item	Type of cost	Total cost per nursery over 3 years	Total cost per child per year over 3 years (34 children per nursery per year)
Programme delivery (NDNA)	Start-up cost per nursery	£2,546.19	£24.96
Optional cover	Start-up cost per nursery	£104.25	£1.02
Optional materials	Start-up cost per nursery	£48.36	£0.47
Optional materials	Recurring cost per nursery	£198	£1.94
Total		**£2,896.80**	**£28.39**

Conclusion

Table 29: Key conclusions

Key conclusions	
1.	Children in nurseries allocated to the intervention group made, on average, the equivalent of three months' additional progress in maths attainment compared to children in control nurseries. This result has a very high security rating.
2.	Children in nurseries allocated to the intervention group made, on average, the equivalent of three months' additional progress in language attainment compared to children in control nurseries.
3.	Children eligible for Early Years Pupil Premium (EYPP) in the intervention nurseries made, on average, the equivalent of six months' additional progress in maths attainment compared to children eligible for EYPP in control nurseries. These results, while promising, should be treated with more caution than the analysis on all pupils as fewer children were included in this analysis.
4.	The results from the implementation and process evaluation support the majority of the components of the logic model, specifically the training and support, the MC and DMC roles, and the action plans, suggesting these should be maintained in the future.
5.	The implementation and process evaluation found that the commitment of the MC and DMC is crucial to successful implementation and that the roles can be interpreted flexibly depending on the needs of the nursery and the choice of nominated MC and DMC.
6.	Subsequent analysis was conducted using Early Years Foundation Stage Profile (EYFSP) assessment data from the National Pupil Database (NPD), to assess the potential long-term impact of Maths Champions on children's attainment. This exploratory analysis found that children in nurseries allocated to the intervention group made, on average, the equivalent of two months' additional progress in mathematics and language attainment compared to children in control nurseries, as well as being slightly more likely to attain a Good Level of Development (GLD) compared to children in control nurseries (equating to approximately one months' additional progress). Children eligible for EYPP in the intervention nurseries made, on average, the equivalent of six months' additional progress in mathematics attainment compared to children eligible for EYPP in control nurseries (however these results, while promising, should be treated with more caution than the analysis on all pupils as fewer children were included in this analysis). With all these estimates there is some statistical uncertainty, as the impact estimate ranges are wide, and therefore should be treated with caution. Full findings from this exploratory longitudinal analysis can be found in Appendix F[9].

Impact evaluation and IPE integration

This rigorously designed, conducted, and reported cluster randomised controlled trial provides significant evidence from the impact evaluation that the MC programme increases children's mathematical and language attainment (three months' additional progress for each). There is also some evidence of a quantitative interaction between trial arm and child EYPP eligibility for the maths outcome: children eligible for the EYPP in nurseries allocated to the MC programme made, on average, six months' additional progress relative to children eligible for the EYPP in control nurseries.

Nurseries were randomised using minimisation and the nursery and child characteristics were well balanced between trial arms at baseline, except for some small observed differences in the proportion of children with EAL and eligible for FEEE (a larger proportion of these children in the control group than intervention). A source of potential post-randomisation bias—attrition—was minimal in this trial, with every nursery providing post-test data (even the four nurseries that withdrew from the programme provided post-test data). At the level of the child, attrition was very low and similar between the two groups—6.0% and 8.6% for the intervention and control groups, respectively. Child-level characteristics and baseline outcome scores were very similar between the 'as randomised' and 'as analysed' samples. Therefore, there was no evidence that attrition introduced selection bias.

All intervention nurseries were at least minimally engaged with the MC programme and three-quarters demonstrated at least good engagement and fidelity. The IPE analysis results are concordant with the results from the impact evaluation, with the staff who were interviewed for the IPE sharing their experiences of the noticeable improvement in children's progress.

As with most evaluations, the results of the trial are only applicable to the sample that were enrolled and randomised into the trial. Consequently, the results may not apply to nurseries that were not eligible to take part. One inclusion

[9] This analysis was conducted later than the publication of the original report in December 2023, and therefore this conclusion was added in June 2024 (full analysis can be found in Appendix F).

criterion was the need to have a minimum of 15 eligible children, nevertheless, we have *no* evidence that the MC programme would not be suitable for smaller nurseries.

Evidence to support the logic model

The findings of both the impact evaluation and the IPE support the vast majority of the components of the logic model. Below we detail the items of the logic model (in italics) and discuss the evidence base from the impact evaluation and the IPE in support (or not) of each logic model item.

Inputs—training, programme activities, and resources

The results of the impact evaluation and IPE support the 'input' components of the logic model. Suitably qualified MCs and DMCs were nominated in 100% of intervention nurseries, 85% of MCs and DMCs having completed the induction webinar and 85% of MCs having completed two compulsory online courses. MCs and DMCs from the 14 nurseries which contributed to the IPE deemed the content, dosage, coverage, and duration of the induction and online courses to be appropriate and they considered them effective at improving early years practitioners' knowledge and understanding of children's mathematical development. The majority of intervention nurseries developed and continued to review a nursery-specific action plan aimed at improving maths provision. Although this was an optional component, 70% of intervention nurseries tracked the progress of children aged three to four throughout the programme (36% very engaged, 33% partially engaged). Most nurseries engaged with the relevant programme resources considered within the logic model to influence outputs. This included implementing ten mandatory resources (67% very engaged), participation in monthly webinars (52% very engaged), and engaging with monthly one to one support provided by the NDNA (73% of nurseries were deemed 'very engaged'). The results of the impact evaluation CACE analysis indicated a greater benefit of the MC programme than in the intention-to-treat analysis, suggesting that the impact of the intervention was even greater when the nursery complied more fully with the inputs of the logic model.

Outputs—dissemination within nursery, involvement of DMC, implementation of action plans, increased communication with parents and carers about children's mathematical development

The unanimous view from IPE participants from 14 nurseries was that the role of the DMC was helpful—that it enabled the MC to share responsibility of implementing the MC programme within the nurseries. Three 'models' of MC-DMC relationship were observed. There was strong evidence from the IPE that practitioners implemented the action plans within their nurseries in daily practice with varying degrees of flexibility. Although data is limited, IPE interview participants commented that parents and carers reported children using more mathematical language at home.

Possible mediators and moderators—the motivation of MCs and DMCs to participate, their qualification level and teaching experience, the frequency of communication with the NDNA, completion of child tracking, and level of resource use

Frequent communication with the NDNA, the completion of child tracking, and the volume of resource use influenced child outcomes as the CACE analysis demonstrated increased benefit in line with compliance. A requirement of the programme was that MCs and DMCs needed to be qualified to at least level 3. Given that MCs and DMCs in all intervention nurseries fulfilled this criterion, and children subsequently made (statistically significant) improvements to their maths and language attainment, we have evidence to support staff qualification as an integral mediator of outcomes within the logic model.

The IPE highlighted that the key necessary condition for the successful delivery of the training and support was the approachable nature of the delivery team (the NDNA). The IPE also identified other factors that contributed towards successful implementation including (1) staff awareness of the programme within the nursery, (2) MCs and DMCs being committed to motivate to deliver the programme as well as having sufficient resource and time, and (3) continuity of implementation in practice by multiple staff members within a nursery, for example, the MC and DMC. How a practitioner's teaching experience influences the delivery and implementation of the MC programme was not explored as part of the impact evaluation so we are unable comment further on this. However, staff qualification level could be deemed a reasonable proxy for teaching experience.

Outcomes—children's maths (primary) and language (secondary) attainment improves at age four, teachers' perceptions of their confidence and competency in maths increase (secondary), maths and language attainment and Good Levels of Development improve at the end of reception year (secondary), and changes are observed to nurseries' maths learning environment and provision (secondary)

There was strong evidence from the impact evaluation of a benefit of the MC programme on the primary outcome, children's maths attainment, and for the secondary outcome of language attainment (reading and phonological awareness). Child-level attrition was minimal and similar between the intervention and control group. With regard to the MC programme increasing practitioner confidence in teaching maths, the results of the impact evaluation and IPE present similar pictures. The impact evaluation found some evidence that the MC programme improved practitioners' confidence in teaching children maths (as measured using an adaptation of Chen et al.'s (2014) 'Early Math Beliefs and Confidence Survey'), and it was the dominant view of staff within the IPE interview data that the MC programme had improved their confidence in delivering maths activities to children. We could hypothesise that it is the MC programme's focus on increasing practitioner knowledge of predictive areas of focus, and how to teach these and monitor children's progress, which led to the significant improvements in children's attainment within the impact evaluation and some evidence of an increase in practitioners' confidence. The impact of the MC programme on children's longer-term maths and language attainment (end of reception, July 2023) will be collected via the EYFSP and reported in a subsequent report addendum.

Interpretation

The impact results of this trial align to the evidence within the EEF's Teaching and Learning Toolkit, which reports that high quality CPD is effective at improving child attainment within the early years, at a low cost.

This was the second effectiveness trial investigating whether the MC programme worked under everyday conditions in a large number of nurseries to improve children's maths attainment. The first effectiveness trial by Robinson-Smith et al. (2018) found that children in the intervention group made, on average, the equivalent of two additional months' progress in maths (effect size 0.10, 95% CI: -0.13 to 0.33, $p = 0.41$) and language (effect size 0.17, 95% CI: -0.06 to 0.40, $p = 0.15$), in comparison to the control group, although in both cases there is uncertainty around the result with the 95% CI crossing zero. The first effectiveness trial obtained a low security rating of two padlocks. While it was recognised that the trial was well designed, the security of the findings was aggravated by high attrition: 36% of the children who were recruited to the trial did not complete the post-test, mainly due to them leaving participating nurseries before the end of the trial. The first effectiveness trial was ultimately underpowered to detect a statistically significant difference as small as 0.10 as was observed for the maths primary outcome and it offered no evidence that the impact of the intervention was moderated by child EYPP eligibility.

In contrast to Robinson-Smith et al. (2018), this trial found that children in the intervention group made, on average, the equivalent of three months' additional progress in the primary outcome of maths (effect size 0.25, 95% CI: 0.12 to 0.38, $p < 0.001$) and secondary outcome of language (effect size 0.21, 95% CI: 0.08 to 0.35, $p < 0.001$), in comparison to the control group. In addition, there was evidence that the benefit was greater among children eligible for the EYPP than for those ineligible. Children eligible for the EYPP in the intervention group made, on average, the equivalent of six months' additional progress in maths relative to EYPP-eligible children in the control group.

Attrition in this trial was minimal and the trial was sufficiently powered to detect an effect size of 0.20 in the primary outcome. All 134 nurseries randomised were retained within the evaluation (although four nurseries randomised to the intervention withdrew part-way through the trial from delivering the programme), and overall child attrition at post-test was 7.3%, none of which were formal withdrawals. Evidently, child movement between, and removal from, early years providers was far lower in this than in the first MC effectiveness trial. When designing the trial, we hypothesised that the inclusion of SNs may bolster child retention; indeed, the proportion of randomised children who left the nursery prior to, or during, post-testing was lower in SN and maintained nurseries (4%) than in PVI nurseries (5%). Anecdotally, we believe this is partly due to parents and carers being keen for their child to receive nursery provision in a school where the child will also continue to attend reception. We also hypothesise that parents and carers were less likely to move their child to a new nursery during the COVID-19 pandemic. The inclusion of SNs enabled us to understand whether the MC programme could also be effective within this type of nursery. The subgroup analyses show no evidence that the impact of the intervention was moderated by type of nursery, indicating that the MC programme can be effectively delivered and improve children's attainment in maths in both PVI nurseries and SNs. Another factor that led to reduced attrition in this trial (in comparison to the first effectiveness trial) was specifying in the eligibility criteria that children

needed to have completed the ASPECTS baseline assessment (in addition to other criteria detailed earlier in this report). In the first effectiveness trial, 136 pupils had missing baseline data, which contributed to the high levels of attrition (Robinson-Smith et al., 2018).

The results of this trial are of particular pertinence given that it was delivered against the backdrop of the COVID-19 pandemic, which saw unprecedented challenges for the early years sector. The implications of the pandemic on young children are noted to have caused delays to communication and language development and PSED (Ofsted, 2022). As a result, we thought it plausible that baseline scores (which were assessed in autumn 2021) could be lower for children in this trial than in the first effectiveness trial, however, this was not the case as ASPECTS baseline maths and language scores were broadly similar between the two trials (in fact, they were slightly higher in the current trial compared to the former; Table 30). This is an interesting finding and may suggest that the maths and language skills of children three years old who attended nursery were not particularly impacted by COVID-19 restrictions. At post-test, however, the trend is reversed as, while scores were still similar between the trials, the ASPECTS maths scores were slightly lower in this current trial than in the first effectiveness trial. This may be a chance finding or is perhaps an artefact of the reasonably high level of attrition seen in the first trial (baseline maths scores of the population 'as analysed' were slightly higher overall than the 'as randomised' population for the first trial). Language scores were very similar between the two trials at post-test. The observed spillover effects of the intervention into language were to be expected due to the implied causal relationship between language and mathematical developed (Chow et al., 2021): children learn mathematics via verbal instruction. Therefore, as practitioners' language and explanations of new mathematical concepts improved in complexity and frequency due to use of the MC programme, so did children's language attainment. Many nurseries that participated in interviews reported that they found the training in the use of mathematical terminology particularly useful and that they had observed an increase in children's use of mathematical language since implementing the MC programme.

Table 30: Comparison of baseline and post-test ASPECTS maths and language scores between the 2016/2017 and 2021/2022 cohorts by randomised group

Assessment	Cohort	Intervention		Control	
		n	Mean (SD)	n	Mean (SD)
ASPECTS baseline maths score	2016/2017 cohort	407	11.6 (6.7)	438	11.0 (6.4)
	2021/2022 cohort	638	12.6 (6.9)	666	12.8 (6.6)
ASPECTS post-test maths score	2016/2017 cohort	304	18.5 (6.2)	349	17.3 (6.2)
	2021/2022 cohort	600	17.9 (6.0)	609	16.5 (6.5)
ASPECTS baseline language score	2016/2017 cohort	407	23.6 (8.7)	438	23.1 (8.2)
	2021/2022 cohort	638	24.4 (9.1)	666	24.6 (8.3)
ASPECTS post-test language score	2016/2017 cohort	304	31.6 (10.3)	392	29.7 (9.8)
	2021/2022 cohort	600	31.7 (9.8)	609	30.0 (9.3)

In September 2017 there was a significant policy change to FEEE which saw the extension of funded childcare from 15 to 30 hours per week (term-time only) for all eligible three- and four-year-olds. In line with this policy change, we postulated at the start of this trial that children's weekly attendance within nursery may have increased in comparison to the average attendance of the 2016/2017 cohort within the first MC effectiveness trial (Robinson-Smith et al., 2018) and consequently that children would have greater exposure to the intervention. However, this policy change did not increase the 2021/2022 cohort's weekly attendance, which bears resemblance to the 2016/2017 cohort's average weekly attendance (see Table 31); both were, on average, approximately 24 hours. Substantially more attendance data was collected as part of this trial in comparison to the Robinson-Smith et al. (2018) trial.

Table 31: Comparison of hours attendance at nursery each week between the 2016/2017 and 2021/2022 cohorts—by randomised group, as included in primary analyses

	Intervention		Control	
	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
2016/2017 cohort	112 (333)	25.1 (11.4)	154 (327)	23.7 (11.3)
2021/2022 cohort	600 (0)	24.6 (8.8)	609 (0)	24.5 (9.3)

In response to the results of the first effectiveness trial of the MC programme (Robinson-Smith et al., 2018), the delivery team made a number of significant changes to the programme, including the exclusion of the BKSB (a tool to assess levels of practitioners' maths capabilities and provide them with practical activities to develop their skills), the introduction of a DMC, and a move from face to face to online induction. All these changes essentially aimed to address the issue of staff burden reported by Robinson-Smith et al. (2018). The delivery team adapted the programme for this trial to provide practitioners with a more comprehensive understanding of the main areas of early years maths in lieu of removal of the BKSB. The role of the DMC was to enable staff to share the responsibilities of programme delivery with their nurseries and to future-proof the delivery of the programme within the nursery should the MC leave or experience long term absence. The move from face to face training to online enabled the programme to be delivered as intended even throughout the COVID-19 pandemic. Within this trial, 74% of nurseries were defined as being very or partially engaged in all the core aspects of the intervention, in comparison to 54% of intervention nurseries participating in the first effectiveness trial (Robinson-Smith et al., 2018). MCs and DMCs from the 14 nurseries that contributed to the IPE within this trial deemed the content, dosage, coverage, and duration of the induction and online courses to be appropriate, and the introduction of the DMC role was viewed positively. Additionally, for this trial, nurseries had up to one month longer exposure to the intervention in comparison to the first effectiveness trial. In this trial, the average time between first NDNA contact and date of withdrawal or post-testing within the nursery was 7.1 months (SD 1.5, range 0 to 8.6) whereas in Robinson-Smith et al. (2018), nurseries had between five and seven months for programme implementation. Together, our trial data suggests that the programme adaptations enabled nurseries to engage better with the programme overall, which consequently saw children achieve better outcomes at post-test.

Many nursery practitioners have a lack of training in maths provision and do not feel confident in their own maths skills (von Spreckelsen et al., 2019). The impact evaluation found some evidence that the MC programme improved practitioners' confidence in teaching children maths and a dominant view within the IPE interview data was that staff perceived this to be the case.

While effective CPD has been shown to improve practice and child outcomes (Brunsek et al., 2020), it can be difficult for early years providers to know where to invest their (limited) resources given the vast CPD resources available. Traditionally, early years CPD programmes focus heavily on improving literacy and PSED (Brunsek et al., 2020), however, maths development plays an equally important role in children's early development as research shows effective teaching practice in early number concepts is predictive of children's maths and reading attainment at Key Stages 1 and 2 (Sammons et al., 2004; 2008), maths and science attainment at Key Stage 3 (Sammons et al., 2011), and even GCSE results (Sylva et al., 2014). Carruthers and Worthington (2009) argue that a key determinant of effective maths CPD is that teaching staff make a 'concept shift' that sees them 'weave in and out of practice and theory' (p.26), which ultimately impacts on their practice and subsequently children's maths. This trial has shown the MC programme to be a successful mechanism for improving the quality of maths practice and, subsequently, positively improving all children's maths and language attainment. There was evidence that the benefit was greater among children eligible for the EYPP than those ineligible. These findings are consistent with other research which highlights the positive effect quality early years education has on child outcomes, the effects of which are most consistent for disadvantaged children (Laurin et al., 2015). This trial has shown the MC programme to be a cost effective and scalable way to improve child outcomes in PVI and SN nurseries in England.

Limitations and lessons learned

We allowed for post-test data to be collected by an unblinded, independent assessor (a member of the evaluation team) or a practitioner or teacher within some nurseries in order to minimise attrition. While this happened in very few cases—33 children (16 intervention; 17 control) across 22 nurseries—this strategy introduced the possibility of bias (for example, unblinded practitioners might have been inclined to help the children or administer the post-test differently in a way that benefits children in one particular group). The impact of ASPECTS being administered by someone other than a blinded RA was investigated in a sensitivity analysis and results were very similar to the primary analysis, indicating this had little impact on the findings.

Nurseries that participated in the IPE had slightly fewer graduate staff in comparison to all nurseries randomised to the intervention group. We acknowledge the possibility that nurseries that took part in the IPE interviews and focus groups may have been different from those that did not volunteer to take part in these aspects and, consequently, the results should be treated with a measure of caution as there is a possibility of a potential source of bias having been introduced due to the self-selection process. For example, stakeholders in nurseries that were extremely enthusiastic about the MC programme may have been more likely to volunteer to be interviewed and, conversely, those less engaged or those with perceived workload issues may have been less likely to take part. This same issue applies to the ECERS data to an even greater extent due to the increased burden on nurseries to undertake these observations and the very small sample of only three nurseries, which is not representative. Only one of the nurseries observed using the ECERS instrument also participated in an interview or focus group; the other two ECERS nurseries did not agree to participate. Therefore, it has not been possible to match up the interview and focus group data with the ECERS data. It was difficult to provide a full and useful interpretation of the ECERS data due to its potential to identify the few participating nurseries.

We also acknowledge the potential for conflict of interest in the NDNA staff members and so this focus group data is treated with a measure of caution. This has to be balanced with the fact that the NDNA provides a unique and valid perception of the programme and its insights supplement the data from the other stakeholder interviews.

Future research and publications

- Given the positive and significant results of this effectiveness trial, further research should explore if the results can be replicated when programme delivery is scaled-up beyond levels seen within this trial.
- Future research could investigate if and how the MC programme is implemented beyond the first year. It would be useful to understand whether nurseries continue to implement the MC approach and the impact that increased ‘soak time’ has on staff’s teaching practice and, consequently, children’s outcomes. The sustainability of the programme could also be explored.
- The evaluation team aim to publish the main findings of the impact evaluation within a high impact peer reviewed journal.

References

All Party Parliamentary Group for Maths and Numeracy (2014) 'Maths and Numeracy in the Early Years': https://nrich.maths.org/content/id/11441/APPG%20paper%20-%20EYs.pdf

Altman, D. G. and Bland, J. M. (2005) 'Treatment Allocation by Minimisation', *BMJ*, 330 (7495), p. 843. https://doi.org/10.1136/bmj.330.7495.843

Asmussen, K., Law, J., Charlton, J., Acquah, D., Brims, L., Pote, I. and McBride, T. (2018) 'Key Competencies in Early Cognitive Development: Things, People, Numbers and Words', London: Early Intervention Foundation: https://www.eif.org.uk/report/key-competencies-in-early-cognitive-development-things-people-numbers-and-words

Borm, G. F., Fransen, J. and Lemmens, W. A. J. G. (2007) 'A Simple Sample Size Formula for Analysis of Covariance in Randomized Clinical Trials', *Clinical Epidemiology*, 60 (12), pp. 1234–1238. https://doi.org/10.1016/j.jclinepi.2007.02.006

Bound, J., Jaeger, D. A. and Baker, R. M. (1995) 'Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak', *American Statistical Association*, 90 (430), pp. 443–450. https://doi.org/10.2307/2291055.

Brunsek, A., Perlman, M., McMullen, E., Falenchuk, O., Fletcher, B., Nocita, G., Kamkar, N. and Shah, P. S. (2020) 'A Meta-Analysis and Systematic Review of the Associations Between Professional Development of Early Childhood Educators and Children's Outcomes', *Early Childhood Research Quarterly*, 53, pp. 217–248. https://doi.org/10.1016/j.ecresq.2020.03.003.

Carruthers, E. and Worthington, M. (2009) 'An Early Years' CPD Initiative for Mathematics: The Power of Collaborative, "Grassroots" Learning', *Proceedings of the British Society for Research into Learning Mathematics*, 29.

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

Connell, A. M. (2009) 'Employing Complier Average Causal Effect Analytic Methods to Examine Effects of Randomized Encouragement Trials', *Drug and Alcohol Abuse*, 35 (4), pp. 253–259. https://doi.org/10.1080/00952990903005882

DfE (2019a) 'Childcare and Early Years Providers Survey: 2019', London, Department for Education. https://www.gov.uk/government/statistics/childcare-and-early-years-providers-survey-2019

DfE (2019b) 'Education Provision: Children Under 5 Years of Age, January 2019', London: Department for Education. https://www.gov.uk/government/statistics/education-provision-children-under-5-years-of-age-january-2019

DfE (2019c) 'National Curriculum Assessments at Key Stage 2 in England 2019 (revised)', London: Department for Education. https://www.gov.uk/government/statistics/national-curriculum-assessments-key-stage-2-2019-revised

DfE (2022) 'Key Stage 2 Attainment, Academic Year 2021/22', London: Department for Education. https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-2-attainment/2021-22

Department of Health and Social Care (2016) 'Factsheet: Developing a Public Health Outcome Measure for Children Aged 2 – 2½ Using ASQ-3TM': https://www.gov.uk/government/publications/measuring-child-development-at-age-2-to-25-years

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K. and Japel, C. (2007) 'School Readiness and Later Achievement', *Developmental Psychology*, 43 (6), pp. 1428–446. https://doi.org/10.1037/0012-1649.43.6.1428

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

EEF (2018) 'Statistical Analysis Guidance for EEF Evaluations', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

EEF (2019) 'Cost Evaluation Guidance for EEF Evaluations', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/Cost_Evaluation_Guidance_2019.12.11.pdf

EEF (2020) 'Improving Mathematics in the Early Years and Key Stage 1', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/tools/guidance-reports/early-maths/#recommendation-1

Evangelou, M. and Mathers, S. (2018) 'Theory of Change: Maths Champions Intervention', in 'Maths Champions: Additional Appendices', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/maths-champions/

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C. and McDowell, J. (2013) 'Teaching Math to Young Children: A Practice Guide', Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/PracticeGuide/18

Hsieh, H.-F. and Shannon, S. E. (2005) 'Three Approaches to Qualitative Content Analysis', *Qualitative Health Research*, 15 (9), pp. 1277–288. https://doi.org/10.1177/1049732305276687

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Handbook.pdf.

Laurin, J. C., Geoffroy, M.-C., Boivin, M., Japel, C., Raynault, M.-F., Tremblay, R. E. and Côté, S. M. (2015) 'Child Care Services, Socioeconomic Inequalities, and Academic Performance', *Pediatrics*, 136 (6), pp. 1112–124. https://doi.org/10.1542/peds.2015-0419

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B. and Shand, R. (2017) *Economic Evaluation in Education: Cost-Effectiveness and Benefit-Cost Analysis (third edition)*, London: SAGE.

Mullis, I. V. S., Martin, M. O., Foy, P. and Arora, A. (2012) 'TIMSS 2011 International Results in Mathematics', Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2011/international-results-mathematics.html

NDNA (2019) 'NDNA 2018/19 Workforce Survey England', Huddersfield: National Day Nurseries Association. https://www.ndna.org.uk/NDNA/News/Reports_and_surveys/Workforce_survey/nursery_workforce_survey_2019.aspx

NHS Digital and Ofsted (2017) 'Ages and Stages Questionnaire (ASQ-3) Analysis: Children and Young People's Health Services Data Set (CYPHS), Experimental Statistics, October 2016 - March 2017', Health and Social Care Information Centre. https://digital.nhs.uk/data-and-information/publications/statistical/children-and-young-people-s-health-services-monthly-statistics/children-and-young-people-s-health-services-monthly-statistics-england-may-2017-experimental-statistics

Office for Health Improvement and Disparities (2021) 'Health visitor service delivery metrics (Experimental Statistics): 2020/2021 annual data statistical commentary (November 2021). Office for Health Improvement and Disparities: https://www.gov.uk/government/statistics/health-visitor-service-delivery-metrics-experimental-statistics-annual-data

Office for Health Improvement and Disparities (2022) 'Children's Public Health 0 to 5 Years: Interim National Reporting Process for the Universal Health Visiting Service. Full Guidance for Local Authority Members of Staff 2022 to 2023', Office for Health Improvement and Disparities. https://www.gov.uk/government/publications/childrens-public-health-0-to-5-years-national-reporting

Ofsted (2022) 'Education Recovery in Early Years Providers: Spring 2022'. https://www.gov.uk/government/publications/education-recovery-in-early-years-providers-spring-2022/education-recovery-in-early-years-providers-spring-2022

Public Health England (2018) 'Feasibility Study: Developing the Capability for Population Surveillance Using Indicators of Child Development Outcomes Aged 2 to 2 and a Half Years', London: Public Health England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/683601/Feasibility_study_developing_the_capability_for_population_surveillance_using_indicators_of_child_development_outcomes_aged_2_to_2_and_a_half_years.pdf

Public Health England (2020) 'Child Development Outcomes at 2 - 2½ Years (Experimental Statistics): Quarter 3 2019/20 Statistical Commentary (April 2020 release)', London: Public Health England. https://www.gov.uk/government/statistics/child-development-outcomes-at-2-to-2-and-a-half-years-2019-to-2020

Public Health England (2021) 'Children's Public Health 0 to 5 Years: Interim National Reporting Process for the Universal Health Visiting Service. Full Guidance for Local Authority Members of Staff 2021 to 2022', Public Health England. https://www.gov.uk/government/publications/childrens-public-health-0-to-5-years-national-reporting

Public Health England (2022) 'Health Visitor Service Delivery Metrics (Experimental Statistics): Annual Data Statistical Commentary 2019 to 2020', Public Health England. https://www.gov.uk/government/statistics/health-visitor-service-delivery-metrics-experimental-statistics-2019-to-2020-annual-data

Quartagno, M. and Carpenter, J. (2023) 'jomo: A Package for Multilevel Joint Modelling Multiple Imputation'. https://CRAN.R-project.org/package=jomo

Robinson-Smith, L., Fairhurst, C., Stone, G., Bell, K., Elliott, L., Gascoine, L., Hallett, S., Hewitt, C., Hugill, J., Torgerson, C., Torgerson, D., Menzies, V. and Ainsworth, H. (2018) 'Maths Champions: Evaluation Report and Executive Summary', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Maths_champions_evaluation_report.pdf

Robinson-Smith, L., Whiteside, K., Torgerson, C., Qi, X., Fairhurst, C., Elliott, L., Hewitt, C., Joshi, K., Menzies, V., Torgerson, D. and Ainsworth, H. (2022) 'Independent Evaluation of Maths Champions in Nursery Settings to Develop Children's Early Numeracy: A Two-Armed Cluster Randomised Controlled Trial—Evaluation Protocol', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/maths-champions-effectiveness/

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley. https://doi.org/10.1002/9780470316696.fmatter.

Rutterford, C., Copas, A. and Eldridge, S. (2015) 'Methods for Sample Size Determination in Cluster Randomized Trials', *International Journal of Epidemiology*, 44 (3), pp. 1051–067. https://doi.org/10.1093/ije/dyv113

Saghaei, M. and Saghaei, S. (2011) 'Implementation of an Open-Source Customizable Minimization Program for Allocation of Patients to Parallel Groups in Clinical Trials', *Biomedical Science and Engineering*, 4 (11), pp. 734–739. https://doi.org/10.4236/jbise.2011.411090

Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., Elliot, K. and Marsh, A. (2004) 'The Effective Provision of Pre-School Education (EPPE) Project: The Continuing Effects of Pre-School Education at Age 7 Years', London: Institute of Education. https://discovery.ucl.ac.uk/id/eprint/10005290/

Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B. and Hunt, S. (2008) 'Effective Pre-school and Primary Education 3-11 Project (EPPE 3-11): Influences on Children's Attainment and Progress in Key Stage 2: Cognitive Outcomes in Year 6', Nottingham: Department for Children, Schools and Families.

Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., Toth, K., Draghici, D. and Smees, R. (2011) 'Effective Pre-school, Primary and Secondary Education Project (EPPSE 3-14): Influences on Students' Attainment and Progress in Key Stage 3: Academic Outcomes in English, Maths and Science in Year 9, Full Report', London: Department for Education.

Senn, S. (1994) 'Testing for Baseline Balance in Clinical Trials', *Statistics in Medicine*, 13 (17), pp. 1715–726. https://doi.org/10.1002/sim.4780131703

von Spreckelsen, M., Dove, E., Coolen, I., Mills, A., Dowker, A., Sylva, K., Ansari, D., Merkley, R., Murphy, V. and Scerif, G. (2019) 'Let's Talk About Maths: The Role of Observed "Maths-Talk" and Maths Provisions in Preschoolers' Numeracy', *Mind, Brain, and Education* (early view article). https://doi.org/10.1111/mbe.12221

Squires, J. and Bricker, D. (2009) *Ages and Stages Questionnaire, Third Edition (ASQ-3)*, Baltimore, MD: Paul H. Brookes.

Stock, J. H., Wright, J. H. and Yogo, M. (2002) 'A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments', *Business and Economic Statistics*, 20 (4), pp. 518–529. https://doi.org/10.1198/073500102288618658.

Sylva, K., Melhuish, E., Sammons, P., Siraj, I., Taggart, B., Smees, R., Toth, K., Welcomme, W. and Hollingworth, K. (2014) 'Students' Educational and Developmental Outcomes at Age 16: Effective Pre-school, Primary and Secondary Education (EPPSE 3-16) Project Research Report', London: Department for Education.

Watts, T. W., Duncan, G. J., Siegler, R. S. and Davis-Kean, P. E. (2014) 'What's Past is Prologue: Relations Between Early Mathematics Knowledge and High School Achievement', *Educational Researcher*, 43 (7), pp. 352–360. https://doi.org/10.3102/0013189X14553660.

Appendix A: EEF cost rating

Appendix Table 1: Cost Rating

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

Appendix B: Security classification of trial findings

Appendix Table 1: Padlock assignment

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	MDES	Attrition					
5	Randomised design	<= 0.2	0-10%	5				5
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%			Adjustment for threats to internal validity **[0]**		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	Low risk of bias due to randomisation and low levels of attrition. Two minor points to note: i) slightly higher attrition in the control arm, ii) use of unblinded assessors on a very small sample of children (33).
Threat 2: Concurrent interventions	Low	There is some evidence of maths related CPD in the control group, but this does not seem to be extensive and there is no evidence to suggest that control schools were induced to take up new interventions in response to their allocation.
Threat 3: Experimental effects	Low	Successful randomisation, with equivalence at baseline. However, there are slightly higher number of EAL and FEEE settings in control group (i.e., control group appears to be more disadvantaged than the intervention group).
Threat 4: Implementation fidelity	Low	Fidelity was high and well monitored.
Threat 5: Missing data	Low	Low levels of missing data (7.3%). Two covariates that were correlated with missingness were included in the adjusted analysis and in the multiple imputations performed. Researchers found to similar results.

Threat 6: Measurement of outcomes	Low	Sound measurement and reporting.
Threat 7: Selective reporting	Low	Well-written and comprehensive report following pre-specified protocol and statistical analysis plan.

- **Initial padlock score:** 5 padlocks
- **Reason for adjustment for threats to validity**: 0 padlocks
- **Final padlock score:** 5 padlocks

Appendix C: Changes since the previous evaluation[10]

Appendix Table 2: Changes since the previous evaluation[11]

Feature		Effectiveness trial I	Effectiveness trial II
Intervention	Intervention content	Completion of BKSB by practitioners One MC at each nursery	No BKSB One MC and one DMC at each nursery
	Delivery model	Face-to-face induction	Online induction
	Intervention duration	It was planned that nurseries would implement the MC programme for 6-7 months	Nurseries were supported to implement the MC programme for approximately 7-8 months to enable the inclusion of school-based nurseries (SN)
Evaluation	Eligibility criteria	Nursery level: PVI nurseries located in Local Authorities in areas of high deprivation (although recruitment was extended to other Local Authorities) Practitioner level: Requirement for a graduate practitioner to be the nominated MC	Nursery level: PVI, maintained nursery schools or children's centres, and government funded infant or primary SN classes (no requirement for nurseries to be from deprived Local Authorities). Requirement for nurseries to have a minimum of 15 eligible children in the recruitment cohort. Requirement for nurseries to not currently be taking part in the evaluation of the Department for Education's Early Years Professional Development Programme. Practitioner level: Practitioners qualified to at least Level 3 (A-level/NVQ Level 3 or equivalent) could be the nominated MC in the absence of a graduate practitioner
	Outcomes and baseline	Nursery environment provision was measured at post-test using the Early Childhood Environmental Rating Scales-	Nursery environment provision, measured pre- and post-intervention using the ECERS-3 and ECERS-E, was

[10] Please delete this section if it is not applicable.

[11] Delete columns from the table if they are not applicable or adjust titles as relevant.

		III (ECERS-3) and the Early Childhood Environmental Rating scale extension (ECERS-E) in all nurseries as a secondary outcome Practitioner Confidence and Beliefs was measured at post-test using an adapted version of Chen et al.'s (2014) 'Early Math Beliefs and Confidence Survey'.	only used to assess a sample of three intervention nurseries for the IPE and this was not a secondary outcome for the impact evaluation. Practitioner Confidence was measured at post-test using subscale 2 only of the adapted version of Chen et al.'s (2014) 'Early Math Beliefs and Confidence Survey'.
	Control condition	Business as usual plus £500 following the completion of post-testing.	Business as usual plus £250 after parent/carer recruitment prior to baseline assessments and £250 following the completion of post-testing.

Appendix D: Effect size estimation

Appendix Table 3: 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**
ASPECTS maths score (primary)	1.43 (0.72, 2.14)	1.58 (0.75, 2.42)	600 (38)	36.5	609 (57)	42.1	39.8
ASPECTS maths score (EYPP subgroup)	4.14 (2.10, 6.17)	3.08 (1.33, 4.82)	66 (6)	47.6	88 (11)	34.2	43.4
ASPECTS language score (secondary)	1.68 (0.60, 2.76)	2.06 (0.73, 3.39)	600 (38)	96.6	609 (57)	86.9	92.3

Appendix E: Maths Champions pilot evaluation report

Background

The purpose of this pilot study was to allow the National Day Nursery Association (NDNA) to pilot changes to the Maths Champions (MC) programme since the first effectiveness trial (Robinson-Smith et al., 2018). Additionally, it gave the evaluation team the opportunity to gauge the usefulness and acceptability of these changes, and to pilot recruitment and retention strategies, and research methods ahead of the second effectiveness trial. However, the pilot study was heavily disrupted by COVID-19. We present the findings of the pilot report here.

Intervention

A detailed description of the intervention and the logic model is provided on pages 10-16 of this document.

Pilot evaluation objectives

Research objective 1: Explore the most efficient way to deliver the MC trial within SN nurseries.

The first Maths Champions effectiveness trial recruited PVI nurseries only. Within the pilot we aimed to determine the most efficient timeline for recruiting school nurseries (SN) and any changes to trial processes that may be required with the inclusion of SN.

Research objective 2: Understand if strategies to gain parental consent from children aged 3-4 are practical and effective.

Aware of the challenges of recruiting children into research trials within the early years, the evaluation team aimed to adopt strategies to maximise the number of consented children per nursery in comparison to previous trials. This was to be achieved by requesting nurseries to provide the total number of children were eligible to participate, after which pre-made information packs were to be sent to the nursery for distribution to each parent/carer. Furthermore, the evaluation team planned to distribute an anonymous survey to parent/carers to complete and return to the nursery, which gather perceived facilitators and barriers to recruitment; the purpose of which was to inform recruitment strategies and materials within the main effectiveness trial.

Research objective 3: Explore if the intended strategy to reduce attrition is practical, feasible and cost effective.

The first Maths Champions effectiveness trial suffered high attrition as many children (19%) left the nursery in-between pre- and post-testing. This pilot aimed to mitigate this by gaining consent from parent/carers at the outset to provide the evaluation team with their child's new nursery destination should they leave before post-testing and conduct post-testing at the child's new nursery or other location where possible. Parents would be contacted prior to the post-testing period to ascertain new nursery destination data, if not already provided by the nursery. Post-testing would be completed by a research assistant, and the new nursery would receive £100 for accommodating the assessment visit.

Research objective 4: Explore the feasibility of recruiting and assessing a 2-3 year old cohort.

The evaluation team would gain parent/carer consent to assess a 2-year old cohort in order to calculate a correlation between the Ages and Stages Questionnaire (ASQ-3) and Assessment Profile on Entry for Children and Toddlers (ASPECTS), so the correlation to be utilised within future early years trials. The eligibility and recruitment process for the 2-year old cohort is described below. It was intended that the ASQ-3 would be completed by nursery practitioners - preferably a child's 'key worker' when the child is 2 years old. Nurseries would be provided with ASQ-3 training and materials by the evaluation team, and completed ASQ-3 questionnaires will be returned via courier arranged by the evaluation team. Children would complete outcome testing with a research assistant in their existing or new nursery (as detailed above) using ASPECTS (described further in the main trial Outcome measures section) in February 2021. The recruitment of a 2-year old cohort was specific to this pilot study and was not intended to be replicated within the main effectiveness trial.

Research objective 5: To explore changes made to the MC programme since the first effectiveness trial and the usefulness and acceptability of these changes within nurseries.

Pilot IPE research questions

The pilot IPE research questions are as follows:

Research Question (RQ) 1: What is the feasibility of evaluating MC within private, voluntary, and independent (PVI) nurseries and SN?

1.1. Is the intended timeline for recruiting PVI nurseries and SN feasible?
1.2. Are intended strategies to improve nursery and child recruitment practical?
1.3. Is the intended strategy to reduce attrition practical and feasible?
1.4. Are intended outcome measures for pre- and post-tests effective and appropriate in terms of cost, administration and evaluation?
1.5. Is the content of baseline and endpoint surveys suitable for capturing necessary data?

RQ 2: What are the barriers to evaluating the MC programme in the pilot?

2.1. What are the barriers to recruiting children in the pilot for pre- and post-tests?
2.2. How has COVID-19 impacted on the proposed delivery of the evaluation methods and what might the impact be of this for the effectiveness trial?

RQ 3: What is the feasibility of delivering MC within PVI nurseries and SN?

3.1. To what extent can NDNA deliver the MC programme and the support to MCs and DMCs as intended in the time allotted?

RQ 4: Is the MC programme implemented with fidelity within PVI nurseries and SN?

4.1. Are nominated staff (Maths Champions and Deputy Maths Champions; MCs, DMCs) accessing the available E-learning modules and the support as specified in the programme plan?
4.2. How effective and appropriate are the level of support and training (e.g. content, coverage, dosage and duration) for MCs and DMCs?
4.3. How is the MC programme disseminated within the nurseries to other staff?
4.4. To what extent do the MCs, DMCs and practitioners implement the MC programme into classroom practice?

RQ 5: What are the different stakeholders' viewpoints on the MC programme?

5.1. What are the perceived impacts of MC?
5.2. What is the perceived role of DMCs?
5.3. What are the perceived impacts of DMCs?

RQ 6: What are the barriers to delivering the MC programme in the pilot?

6.1. What are the barriers for MCs and DMCs to engage with induction and E-learning modules?
6.2. What are the barriers for MCs, DMCs and practitioners to implement MC in their classroom practice?
6.3. What are the barriers for MCs to disseminate MC to other staff in the classroom?

RQ 7: What appear to be the necessary conditions for the successful delivery of MC programme?

7.1. What are the necessary conditions for successful recruitment of nurseries, families and children?
7.2. What are the necessary conditions for MCs and DMCs to engage with the training and E-learning modules and the monthly one-to-one support?
7.3. What are the necessary conditions for practitioners to implement MC into practice?
7.4. What are the necessary conditions for reducing the attrition in the pilot trial?

RQ 8: How could the MC programme be improved?

RQ 9: How could the delivery of the MC programme be improved?

The following RQs were added to the pilot IPE in response to the COVID-19 pandemic and subsequent amendments to protocol version 2.0 (Robinson-Smith et al. November 2020).

RQ 2.2. How has COVID-19 impacted on the proposed delivery of the evaluation methods and what might the impact be of this for the effectiveness trial?

RQ 10: How has the delivery of the MC programme been impacted by the COVID-19 pandemic?

Ethics and trial registration

The Health Sciences Research Governance Committee at the University of York granted ethical approval for the pilot study on 29th November 2019. The School of Education Ethics Committee at Durham University also gave full ethical approval.

Data protection

As noted on page 18 of the main effectiveness trial report.

Project team

As listed on page 19 of the main effectiveness trial report.

Methods

Design

The pilot study adopted a non-randomised study design where all participating nurseries received the Maths Champions programme. Two cohorts of children were recruited into the pilot:

- Cohort 1: 3-4 year olds
- Cohort 2: 2-year olds

The pilot study commenced in January 2020, as to inform the main effectiveness trial which was due to start in September 2020, with post-testing for the 2-year old cohort scheduled for February 2021.

Participant selection

Nurseries

Recruitment of nurseries (target n = 12) for the pilot was led by NDNA. Strategies for recruitment included emails to nurseries and advertisements (including on NDNA's website).

Nursery eligibility criteria were:

- PVI providers based on non-domestic premises, maintained nursery schools or children's centres, or government funded infant or primary SN providing nursery provision for 3 and 4 year olds (who were due to begin reception in September 2020).
- Nurseries who had a minimum of 15 children in the cohort starting reception in September 2020.
- Nurseries that were not currently using the NDNA MC programme and had not done so in the past.
- Nurseries that were not taking part in the evaluation of the Department for Education's Early Years Professional Development Programme.
- Nurseries that agreed to all requirements outlined in the Information for Nurseries and Memorandum of Understanding (MoU) document.

Nurseries that were willing to participate returned a completed and signed MoU. Nurseries were also required to sign a Data Sharing Agreement (DSA) issued by the University of York, and an End User Licence Agreement (EULA) issued by CEM at Cambridge Assessment for use of the ASPECTS assessment.

Nurseries participating in the pilot received a thank you payment of £250 following parent/carer recruitment, prior to baseline assessments, and a further £250 after outcome testing.

Children

Eligibility criteria for the two cohorts of children recruited within the pilot are listed below.

Cohort 1:

- Children, aged 3 to 4 years, due to start reception class in school in September 2020.
- Children who attended nursery for a minimum of 15 hours per week.
- Children whose parents/carers anticipated they would remain at the nursery for the duration of the pilot (until June 2020).

Cohort 2:

- Children aged 2 by 1st January 2020 or aged 3, due to move to reception class in September 2021 or September 2022.
- Children who attended nursery for at least 15 hours per week.

- Children whose parents/carers anticipated they would remain at the nursery for the duration of the pilot (until February 2021).

Nurseries were provided with paper information sheets and consent forms to distribute to the parents/carers of all eligible children for each cohort. Nurseries were not required to recruit a 2–3-year-old cohort to participate in the pilot, but this was encouraged. Parent/carers returned completed paper consent forms to nurseries and these were returned to YTU via courier. Nurseries completed a password-protected spreadsheet with relevant participating child details.

The target number of children to be recruited per nursery per cohort is noted below:

- Cohort 1: 10 children
- Cohort 2: no target

In nurseries that gained parent/carer consent for more than 10 consent forms for Cohort 1 children, the evaluation team randomly selected 10 children to complete ASPECTS.

Outcome measures

Cohort 1

ASPECTS is detailed on page 25 of the main effectiveness trial report. It was planned for ASPECTS to be administered either by a nursery practitioner, who was provided with a 15-minute training webinar and written instructions on how to complete ASPECTS, or a trained research assistant from the evaluation team (for school nurseries that requested a research assistant to visit rather than completing ASPECTS themselves). Research assistants would visit each nursery to complete outcome assessments with the same children.

Cohort 2

The ASQ-3 (Squires and Bricker, 2009) captures communication, gross motor, fine motor, problem solving, and personal-social development of babies and young children aged 1 to 66 months. There are 21 versions of the questionnaire, each suitable for a certain age range. It can be completed by parents/carers or other practitioners, including nursery practitioners.

ASQ-3 is one of the limited number of assessments that can be conducted by early years practitioners in nurseries to capture development in young children (some nurseries currently routinely use this assessment). The ASQ-3 is used as part of the Healthy Child Programme health and developmental review that all children in England should receive between 2-2.5 years, the results of which are logged on an NHS Digital database (Department of Health and Social Care, 2016; Public Health England, 2018; 2020).

It was planned that nursery practitioners would complete the ASQ-3 for the Cohort of 2-3 year old children in February 2020 and then the same children would complete ASPECTS with a research assistant in February 2021.

Statistical analysis

Cohort 1

The purpose of conducting baseline and post-testing with this cohort was to pilot processes and procedures, as detailed above, prior to implementation within the effectiveness trial. Therefore, no formal analysis of ASPECTS was to be undertaken for Cohort 1; descriptive data are provided relating to baseline and outcome assessment completion rates.

Cohort 2

ASQ-3 scores at 2-3 years old and ASPECTS scores at 3-4 years old would be summarised descriptively and the Pearson’s correlation between the two will be presented with a 95% confidence interval.

Implementation and process evaluation design

The pilot implementation and process evaluation (IPE) was designed to ensure adherence to the key principles for the design, conduct and reporting of the impact evaluation. The pilot aimed to address the descriptive and experiential aspects of the pilot research questions, listed above. The cross-sectional design will explore the perceptions and experiences of key stakeholders at the beginning of the pilot to provide snapshot descriptive data on perceptions about recruitment and towards the end of the pilot study period to provide snapshot descriptive data on perceptions about: barriers and facilitators to recruitment and retention; feasibility and acceptability of MC implementation and delivery; and acceptability and feasibility of undertaking the baseline and outcome assessments.

Research methods

The pilot study aimed to recruit 4/12 participating nurseries (two PVI, two SN) which would form the sample for the pilot IPE. Nurseries were emailed participation invites and online consent was sought. Data collection methods included interviews and focus groups with key stakeholders within nurseries, and NDNA. Usual practice surveys were also developed. Table 1 provides an overview of the IPE methods for the pilot study.

Table 1: IPE pilot methods overview

Feature		Research Method	Data collection methods	Participants/ data sources (type, number)	Research questions addressed	Implementation/ logic model relevance
Recruitment Delivering MC to PVI and SN		Cross-sectional	Semi-structured interview/ focus group	NDNA staff (n = 2)	RQ 1: 1.1; RQ 3: 3.1; RQ 4: 4.1 RQ 6: 6.1; RQ 7: 7.1; RQ 8; RQ 9	Feasibility; Fidelity; Context
Recruitment Delivering MC to PVI and SN		Cross-sectional	Semi-structured interview/ focus group	MCs (PVI n = 2; SN n = 2)	RQ 4: 4.2, 4.3, 4.4; RQ 6: 6.1, 6.2; RQ 7: 7.2; RQ 8; RQ 9	Fidelity; Context
Recruitment Delivering MC to PVI and SN		Cross-sectional	Semi-structured interview/ focus group	DMCs (PVI n = 2; SN n = 2)	RQ 4: 4.2, 4.3, 4.4; RQ 5: 5.1, 5.2; RQ 6: 6.1, 6.2; RQ 7: 7.2; RQ 8; RQ 9	Fidelity; Context
Recruitment Delivering MC to PVI and SN		Cross-sectional	Semi-structured interview/ focus group	Other practitioners (PVI n = 2; SN n = 2)	RQ 4: 4.3, 4.4; RQ 5: 5.2; RQ 6: 6.2; RQ 7: 7.3; RQ 8; RQ 9	Fidelity; Context
Role of DMC and other MC content change(s)		Cross-sectional	Semi-structured Interview/ focus group	MCs (PVI n = 2; SN n = 2)	RQ 5: 5.2, 5.3	Process outcomes (confidence and competence)
		Cross-sectional	Semi-structured interview/ focus group	DMCs (PVI n = 2; SN n = 2)	RQ 5: 5.2	Process outcomes (confidence and competence)
		Cross-sectional	Semi-structured interview/ focus group	Other practitioners (PVI n = 2; SN n = 2)	RQ 5: 5.2, 5.3	Process outcomes (confidence and competence)

Pilot trial data collection processes		Cross-sectional	Semi-structured interview/ focus group	Sample staff (n = 4)	RQ 1: 1.4; RQ 4: 4.4; RQ 6: 6.2; RQ 2: 2.1; RQ 7: 7.3	Feasibility; Fidelity
				NDNA staff (n = 2)	RQ 5: 5.1, 5.3	Compliance;
		Longitudinal design	E-learning logs data	All (n = 12)	RQ 1: 1.2, 1.3, 1.4; RQ 2: 2.1; RQ 7: 7.4	Compliance; Context
Parental perceptions		Cross-sectional survey	Paper or Online surveys	Parents of eligible children in all nurseries (n = 12) are invited to complete the survey anonymously	RQ 1: 1.2, 1.3; RQ 2: 2.1; RQ 7: 7.1	Context; Feasibility
Baseline and end-point usual practice surveys		Cross-sectional (start and end of pilot)	Paper or online	Manager/head teacher in all control and intervention nurseries	RQ1.5	Feasibility; context
COVID-19		Cross-sectional	Semi-structure interview/ focus groups	All participants in all interviews/focus groups	RQ 10 RQ 2.2	Perceived impact of COVID-19 pandemic

Analysis

IPE data was analysed using a combination of inductive and deductive analyses. Emerging patterns in the data were grouped thematically according to the research questions. Results were synthesised from the themes and presented as answers to each pilot IPE research question.

Timeline

As shown in the timeline in Table 2, the COVID-19 pandemic and lockdowns heavily impacted on the pilot study timeline and evaluation activities.

Table 2: Pilot study timeline, displaying the disruption caused by COVID-19

Original Dates	Completed Dates	Activity	Staff responsible / leading
Nov - Dec 2019	Nov - Dec 2019	Recruit nurseries	DT (support from ET)
Jan 2020	Jan - Feb 2020	Recruit parent/carers of children for cohort 1 and cohort 2	ET
Jan - Feb 2020	Jan - Mar 2020 (baseline assessments with cohort 2 children suspended due to COVID-19)	Pilot baseline assessments with children in cohort 1 and 2; nursery usual practice survey	ET
Feb 2020	Mar 2020 then paused due to COVID-19	Pilot nurseries commence MC programme (support and resources provided for 12 months)	DT
Jan - Jun 2020	Apr 2020	IPE interviews with NDNA	ET
Jun 2020	Activity suspended due to COVID-19	Pilot outcome assessments with cohort 1 children; IPE end-point nursery/staff surveys	ET

Original Dates	Completed Dates	Activity	Staff responsible / leading
Aug - Sep 2020	Sept 2020	Submission of pilot study interim report	ET
-	Oct - Nov 2020	Restart delivery of MC programme to pilot nurseries (support and resources provided for 12 months)	DT
-	May - Jun 2021	Collect current nursey destination for cohort 2 children	ET
Feb 2021	Activity suspended due to COVID-19	Outcome assessments with cohort 2 children	ET
Jan - Jun 2020	Jan - Feb 2021	IPE interviews with nurseries	ET

Note: ET = Evaluation Team; DT = Delivery Team

Summary of impact of COVID-19 on evaluation activities

Nurseries completed the MC programme induction in early March 2020, but programme delivery and evaluation activities with nurseries was subsequently paused shortly after due to COVID-19 lockdowns. Activity resumed in October 2021 where participating nurseries received an additional induction webinar. Nurseries subsequently had access to the Maths Champions programme for 12 months, until September 2021. Consequently, there were a number of changes to the pilot study:

- Parents/carers of Cohort 1 children were not asked to complete a questionnaire to gather perceived facilitators and barriers to recruitment.
- Following baseline data collection, Cohort 1 did not complete outcome testing in June 2020 as planned as they had left the nursery to start school by the time the pilot resumed in October 2020.
- Practitioners were not asked to complete the ASQ-3 with Cohort 2 (children aged 2-3 years old). As outlined in the revised protocol (Robinson-Smith et al., 2022), the evaluation team planned to ask nurseries if they routinely completed ASQ-3 and, if so, to provide any ASQ-3 data they already have for participating 2-3 year olds, to gain information about routine use of the ASQ-3 and the feasibility of collecting this data from nurseries. However, due to the (omicron variant) COVID-19 outbreak between November 2021 - February 2022, evaluation activities over this period were paused again and consequently pilot nurseries were not asked if they held ASQ-3 data.
- A revision was made to the protocol, stating that the outcome assessments for Cohort 2 planned in February 2021 were contingent upon the COVID-19 situation. The purpose of these assessment was also revised; instead of providing ASPECTS scores at 3-4 years to correlate with ASQ-3 scores provided by practitioners at 2-3 years, the intention was to explore the feasibility of completing ASPECTS in children's current or new nursery. This would have offered the opportunity to gauge nurseries' perspectives on the COVID-19 protective measures necessary to enable assessment visits and would have allowed the evaluation team to pilot the strategy for locating and assessing children who have moved nursery, which would be useful for reducing attrition in the main effectiveness trial (a process originally planned to take place with children in Cohort 1 of the pilot study). Nevertheless, due to the (omicron variant) COVID-19 outbreak between November 2021 and February 2022, it was not possible to carry out these assessment visits. As planned, nurseries were contacted to collect information on child destination to assess levels of attrition.

Results

This section presents a summary of pilot recruitment and data collection activities.

Recruitment

Nurseries

Between January-March 2020, 11 nurseries participated in the pilot study. The delivery team initially recruited 12 pilot nurseries (all of whom returned a signed MoU); however, one nursery was excluded before the child recruitment stage, as agreed with EEF, due to being based within a private school.

Children

Cohort 1 (3-4 years old)

Participating nurseries reported that 453 children met all the eligibility criteria (median 30; range 19-119). Following the distribution of parent/carer information sheets and consent forms, a total of 301 (66% of those eligible) children were recruited (median 18; range 9-96); 10/11 nurseries recruited the minimum of 10 children. A further 16 consent forms were received although deemed to be invalid (e.g., consent form not being signed or the consent statements not being ticked/initialed; children too young).

Nurseries were initially requested to collect completed consent forms from parents/carers, who were happy for their child to take part, over a period of a week and a half. In practice, consent forms were received at YTU from all nurseries within 3 weeks following the date they were asked to distribute the parent/carer information by.

Cohort 2 (2-3 years old)

In total, 10/11 nurseries reported having 192 children meeting the relevant eligibility criteria (median 18.5; range 3-45) however only 9 nurseries recruited a Cohort 2, with consent forms received for 77 children (40% of those eligible, (median 8; range 3-15). One consent form received was invalid. A small number of additional consent forms for this cohort were also returned but were not entered into the trial database before the evaluation team started working from home due to COVID-19.

Baseline data collection

Child details

Each nursery was asked to compile and securely return a spreadsheet of child details for children who they received a completed parent/carer consent form for; 10/11 pilot nurseries were returned. Baseline details for children recruited into Cohort 1 and Cohort 2 are provided in Table 3, as well as details for the 88 children in Cohort 1 who completed ASPECTS at baseline. Mean age at recruitment is presented in months in Table 3 and is equivalent to 3 years 11 months for children recruited and assessed for Cohort 1 and 2 years 11 months for children recruited for Cohort 2.

The proportion of children reported to be in receipt of funded childcare at 2 years old (free early education entitlements, FEEE) is much lower in Cohort 1 than Cohort 2.

Pupil eligibility criteria for Cohort 1 stipulated that children must attend the nursery for a minimum of 15 hours per week to be eligible to participate in the evaluation. Table 3 details that the average weekly attendance for Cohort 1 was 21.3 hours (range: 11-50 hours), which is similar to the average weekly attendance documented in Robinson-Smith (2018).

Table 3: Baseline characteristics for children recruited in the pilot study and for the children who completed ASPECTS

Child-level (categorical)	Cohort 1 (3–4-year-olds)				Cohort 2 (2–3-year-olds)	
	Recruited		Completed ASPECTS at baseline		Recruited	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Gender, Male	162/301 (9)	162 (53.8%)	52/88 (0)	52 (59.1%)	41/77 (6)	41 (53.2%)
Child in receipt of Early Years Pupil Premium (EYPP)[1]	33/301 (55)	33 (11.0%)	7/88 (10)	7 (8.0%)	N/A	N/A
Child received funded childcare (FEEE) when they were 2 years old	20/301 (174)	20 (6.6%)	7/88 (30)	7 (8.0%)	20/77 (21)	20 (26.0%)
Neighborhood deprivation (IDACI)[2] 20% most deprived 20-40% 40-60% 60-80%, plus 20% least deprived	243/301 (58)	 31 (10.3%) 73 (24.3%) 65 (21.6%) 74 (24.6%)	77/88 (11)	 12 (13.6%) 29 (33.0%) 17 (19.3%) 19 (21.5%)	55/77 (22)	 7 (9.1%) 23 (29.9%) 14 (18.2%) 11 (14.3%)
Child-level (continuous)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
Attendance at nursery per week (hours)[3]	292/301 (9)	21.3 (7.6)	88/88 (0)	23.5 (8.1)	71/77 (6)	17.5 (6.9)
Age at recruitment (months)[4]	301/301 (0)	47.0 (3.5)	88/88 (0)	47.1 (3.5)	77/77 (0)	35.1 (5.1)

[1] Children become eligible for EYPP from 3 years old.
[2] Income Deprivation Affecting Children Index (IDACI) scores were obtained using the children's home postcodes (scores retrieved here: http://imd-by-postcode.opendatacommunities.org/imd/2019). England has been split up into 32,844 small areas and these have been ranked according to the proportion of children in each area who live in income deprived families. The most deprived 20% reported here reflects IDACI decile 1 and 2 areas, 20-40% reflects decile 3 and 4 areas, 40-60% reflects decile 5 and 6, 60-80% reflects decile 7 and 8, and the 20% least deprived reflects decile 9 and 10 areas. Deciles 5 and 6 condensed to avoid disclosive cells.
[3] 2 children recruited (but not assessed) for Cohort 1 and 7 children recruited for Cohort 2 attended nursery for less than 15 hours a week and therefore should have been classed as ineligible for the pilot. Child details spreadsheets will be checked prior to ASPECTS selection in the effectiveness trial to ensure recruited and randomised children are reported as attending for at least 15 hours a week.
[4] Age at recruitment was defined as age on the date the consent from was received at YTU.

Cohort 1

ASPECTS baseline assessments were completed at 9/11 nurseries with a total of 88 children in February/March 2020, which is 29.2% of the 301 recruited sample. The remaining nurseries did not complete ASPECTS before the study was paused due to COVID-19.

As per the protocol at the time, SN were asked if they would prefer for a member of the research team to visit and complete ASPECTS or if they would rather carry out ASPECTS themselves. Only one (which self-classified as a maintained nursery) of these four nurseries chose for a member of the research team to visit and complete ASPECTS.

As 10/11 nurseries recruited more than 10 children, a YTU statistician randomly selected 10 children to complete ASPECTS at each of these nurseries, as well as a number of reserve children. Five of the nurseries needed to assess some reserve children instead of randomly selected children, due to selected children being absent on the assessment day(s), or not being able to access ASPECTS due to significant Special Educational Need(s) or having English as an additional language with very little English language competence.

ASPECTS baseline scores for the 88 children that completed the assessment are presented in Table 4. The data were roughly normally distributed, with no floor or ceiling effects.

Table 4: Baseline ASPECTS data for Cohort 1

Child-level (continuous)	n/N (missing)	Mean (SD)
Maths ASPECTS score	88/90 (2)	14.9 (7.0)
Literacy/language ASPECTS score	88/90 (2)	26.8 (10.2)
Age at assessment (months)	88/90 (2)	48.3 (3.4)

Notes: The N provided is the number of ASPECTS assessment requested to be completed in the 9 nurseries that completed ASPECTS. The two nurseries that did not complete ASPECTS are excluded from the N for this table.
Possible maths scores range from 0-29 and possible literacy/language scores (reading and phonological awareness) range from 0-53, with higher scores indicating greater attainment.

It was not possible to collect post-test data from Cohort 1 as the study was paused due to COVID-19, and by the time it had resumed all children had moved from nursery to reception.

Cohort 2

Nurseries who recruited a 2-3 year old cohort were not asked to complete the ASQ-3 as the study was paused due to the COVID-19 pandemic.

In Spring 2021, nurseries were requested to complete a spreadsheet to detail if participating Cohort 2 children were still present at the nursery. This was an attrition monitoring exercise to inform the main effectiveness trial, in lieu of being able to do this for Cohort 1.

Of the 77 children that were recruited to Cohort 2 from 9 nurseries, new destination data was obtained for 68/77 children from 7/9 nurseries. In total, 53/68 children remained at the same nursery and 15/68 children had left. The name of the new nursery was provided for 3/15 leavers from 3 different nurseries. For the remaining 12/15 leavers, the new destination was home or unknown. Given the ongoing implications of the COVID-19 pandemic, it was not deemed appropriate or feasible to request to visit new destinations to assess the participating children.

Implementation and process evaluation results

Baseline survey

The baseline survey was completed by 10/11 pilot nurseries in February/March 2020; 9 were fully completed.

The survey was delivered via Qualtrics survey software and took nurseries a median time of 11.5 minutes to complete (range = 5.7-67.1 minutes) and thus was deemed an acceptable length for the effectiveness trial.

The survey showed that three nurseries had completed math CPD and five nurseries had completed CPD in another subject in the past two years. Eight nurseries deemed their staff would benefit most from maths-focused CPD in comparison to other subjects. Five nurseries reported that they have a certain amount of time allocated for CPD each

week. Four nurseries stated their annual CPD budget (median £3,250; range £1000-£10,000), one nursery said they do not have a budget and it depends on needs at the time, another said all CPD is done in-house.

Across the nurseries, the highest level of early years qualification for DMCs varied between Level 3 to graduate level (Level 6), however the majority of nominated MCs were qualified to graduate level.

Interviews

It was originally intended that the pilot study would recruit staff from 4 participating nurseries (2 SN; 2 PVI) to participate in interviews. Despite all participating nurseries been invited to participate, only 1 MC within 1 nursery agreed and completed an interview; the impact that the pandemic had on staffing levels and staff's time was noted as a primary reason for non-participation. It was agreed with EEF that the evaluation team would introduce a £10 high street gift vouchers to participants as a thank you for taking part in the interviews, in an attempt to increase participation.

The MC showed a very positive overview of the MC programme and found the support from NDNA-the delivery team was very satisfactory. The MC was very satisfied with the programme induction sessions and noted that the optional induction sessions were particularly helpful so that she could attend the session at her convenience. She found that the MC programme materials and resources are very helpful and she partially enjoyed the three courses (e-learning modules).

The MC explained that it has been difficult to work with the DMC at times in implementing the MC programme in the nursery due to the impact of the COVID-19 pandemic. The MC and DMC have been working in different 'bubbles' within the nursery, and therefore they were unable to work together to develop an implementation plan. When the teaching activities returned to 'normal' (i.e., no longer needed to stay in different bubbles), the MC and DMC managed to catch up weekly to discuss the MC programme implementation plan and the MC noted that it's '*very helpful to have a DMC* ', because '*we can share the activity*' in the nursery.

The MC noted that they managed to meet with the practitioner weekly to discuss the MC programme implementation when the situation allows. However, there have been difficulties to disseminate the MC programme and resources to the practitioner in her classroom due to the change of staff during the term time and the disturbance of teaching sessions due to staff and children's absences (on one occasion, the whole class had to be sent back home). In addition to the MC programme, there has been minimum support and resources available to staff in the nursery.

The MC also noted that the workload associated with the MC programme was '*manageable but didn't have much time*'.

The MC noted that for a successful implementation of the MC programme, a supportive nursery environment is very important. She noted that it is also essential to provide all necessary resources and materials to implement the MC programme because '*not every setting has the required resources'*. Access to the programme is also an essential condition, because *'it is not always easy to get access to the (MC) programme when I was working from home*, and while working in the nursery, '*it's difficult to find a quiet place*'.

Delivery team focus groups

Two focus groups were completed with members of the delivery team in April 2020 (first interview had 3 participants; second interview had 2 participants). The main themes emerging from the interviews with the delivery team and recommendations for the pilot study and effectiveness trial are discussed below.

Rationale and programme changes

Interview data indicated that the aim of the Maths Champions programme has changed since the first effectiveness trial (Robinson-Smith et al. 2018) to focus on practitioners' confidence in understanding how children develop in mathematical thinking, rather than on additionally developing confidence in practitioners' own mathematical abilities. It was stated that '*practitioners need to understand how young children learn maths rather than practitioners developing their maths'* (Delivery team interview 2). This was the rationale for the removal of the requirement for practitioners to use the Basic and Key Skill Builder (BKSB; a tool to assess levels of practitioners' math capabilities and provide them with practical activities to develop their skills) for the current evaluation and also the rationale for a change in focus in some of the training for the MCs and DMCs. Within the first evaluation of the Maths Champions programme (Robinson-

Smith et al., 2018), a requirement of the programme was that the MC within each nursery needed to be qualified to at least degree level. Interview data indicated that the delivery team do not have any reservations related to the lowering of MC qualifications, as this change is believed to be more representative of real-world nurseries.

The delivery team also highlighted further changes to the programme to reduce the burden on MCs, including reducing the number of steps within the programme (from 12 to 9) and the introduction of the new DMC role. Another important change was moving the induction from being face-to-face to online, which the delivery team felt was positively received by MCs in general. The delivery team believed the strengths of the online induction included '*the flexible two-time options (lunch-, and evening- sessions)*', '*the screen sharing and easy demonstration process*' and '*the recap of recorded videos in case of missing sessions'* (Delivery team interview 1).

Inclusion of School Nurseries

The delivery team noted that they believed the process of recruiting nurseries for the pilot study went quite smoothly. The delivery team reported no particular issues with the nursery inclusion criterion for the pilot study; however, for the main effectiveness trial, they found that the exclusion criteria relating to participation in the DfE EY PD Programme was a substantial barrier as many nurseries stated they were taking part in that research.

Inclusion of SN nurseries is a significant change to the sample from the previous effectiveness trial which only included PVI nurseries (Robinson-Smith et al., 2018). The delivery team noted (minor/slight) differences between working with PVI and SN. Firstly, it was stated that the recruiting process for PVIs took slightly longer than the process for SN. This is because the delivery team *'used a charity called Early Education to spread the message to schools and the team got an influx of schools quite quickly whereas, with PVI settings, it took a slightly longer time to spread the message through the NDNA's communication*' (Delivery team interview 1). Secondly, it was reported that communication was more of an issue within SN, as it was often more difficult to get hold of the main contact. *'Settings were very hard to get on the phone because they're always in more of a setting situation that is teaching and busy all the time.'* (Delivery team interview 1). Thirdly, in comparison to PVI, SN were more restricted in their process relating to the use of photographs and access to and use of the Maths Champions community page, which is based on the social media platform, Facebook, where practitioners are invited to share their practice through pictures of the activities they have delivered. *'Within schools, they have to get that sign[ed] from the headteacher and that's taking a little bit more time within some settings. For example, some of the [SN] settings said they were not going to join the Facebook page which is an option that the delivery team made available for them.'* (Delivery team interview 1). Nonetheless, from the delivery team's perspective, there were not many differences between working with the two types of nurseries, *'it was surprising to see how little difference has been found working between the two different types of setting'* (Delivery team interview 2).

Barriers to Maths Champions programme delivery

From the delivery team's perspective, there were no barriers to the delivery and continuation of the programme by NDNA within the pilot study (excluding those posed by the pandemic). However, they assumed that from MC's perspective, *'time will be a key barrier to make sure they have time to do something'* (Delivery team interview 2), and that is why they have created the role of DMC to support the MC.

At a nursery-level, the delivery team highlighted a number of factors from their initial experience with pilot nurseries that may be a barrier to progression and continuation of the programme. For instance, the delivery team noted that some pilot nurseries were not able to fully commit to the timeframe for the programme: *'one setting in the pilot trial already fell a little bit behind and (the delivery team) were struggling to get hold of the right people there*' (Delivery team interview 1). This could be related to staffing issues within nurseries, as the delivery team discussed the issue that *'one or two nurseries hadn't filled in step one and that this is due to the DMC having been on paternity leave and was due to return in the following week and they were waiting until he returned'* (Delivery team interview 1). It was also noted that when the pilot restarts, there may be an issue with one pilot nursery that had not logged onto the platform following the induction, before the study was paused due to the COVID-19 pandemic.

The delivery team's other possible issues surrounding the progression and continuation of the programme in nurseries were related to the pandemic. Technology-wise, it was mentioned that '*it's just logistics issue around what's happened with the situation at the moment from doing the programme within a setting to doing it in a home environment'* (Delivery team interview 1). The delivery team was also concerned that time will be the issue as they do not know what nursery provision and practice is going to look like beyond COVID-19.

Necessary/desirable conditions for the success of the Maths Champions programme

Good communication between the delivery team and the nurseries appeared to be necessary and important. It was noted that *'the success of the programme depends on the context of the nurseries. Recruit the right setting and speak to the right person within the setting and communicate with them regularly are all necessary conditions'* (Delivery team interview 1). It was also indicated that time and support from each nursery's management team are necessary conditions.

With regards to the Maths Champions programme itself, the delivery team highlighted that it is important to see the impact of the newly developed mandatory list of the elements of the programme and how practitioners fit these mandatory elements into their practice. It is also necessary that *'for the MCs and DMCs, at the end of the programme they can reflect on their action plan about what they have achieved and what the next steps are; the teaching quality doesn't/should/t stop at the end of the programme, so they need to continue on the journey*' (Delivery team interview 2).

Parent/carer survey

The anonymous parent/carer survey on facilitators and barriers to recruitment was not distributed as planned due to COVID-19 nursery closures.

Conclusions

Research objective 1: Explore the most efficient way to deliver the MC trial within SN nurseries.

The pilot phase of the evaluation of Maths Champions has provided valuable lessons that were implemented in the delivery of the effectiveness trial. The pilot showed that recruiting both PVI and SN was possible and recruitment strategies were effective. Within the pilot trial, a nursery within a private school (non—state funded) was recruited and later withdrawn. Subsequently nursery eligibility criteria were refined and, notably, a nursery eligibility survey will be developed for use in the main effectiveness trial recruitment.

The pilot highlighted complexities in determining the 'type' of nursery, specifically regarding maintained nurseries. In the effectiveness trial, the aim is to recruit 70% PVI nurseries and 30% SN; however, the original protocol did not specify in which category 'maintained' nurseries should be. Maintained nurseries are local authority funded (Early Education, 2015), though there is currently uncertainty regarding their ongoing funding (Early Education, 2018; National Education Union, 2020). Maintained nurseries are legally constituted as schools and have a governing body, at least one practitioner with qualified teacher status, and a head teacher who, unlike in SN, is an Early Years specialist (Paull and Popov, 2019). For the effectiveness trial, nurseries will provide their type on the eligibility survey. For the main effectiveness trial, maintained nurseries will be grouped with SN for minimisation for randomisation and the sensitivity analysis. The 'get-information-schools.service.gov.uk' website will also be used to confirm nursery type.

Research objective 2: Understand if strategies to gain parental consent from children aged 3-4 are practical and effective.

The pilot showed that it was possible for nursery staff to recruit a 3–4-year-old cohort with 66.4% of eligible children being recruited. Some nurseries were very successful in recruiting a large proportion of the cohort.

As detailed earlier in this pilot report, 5/10 nurseries needed to assess 'reserve' children, one reason being because the practitioner deemed the child unable to engage with the assessment due to significant SEND. Eligibility criteria were subsequently added to the protocol to exclude children from the main effectiveness trial, at recruitment, if practitioners consider them to have significant Special Educational Need(s) or Disabilities, or English as an Additional Language where an extreme language barrier exists, which would prevent them from accessing the ASPECTS assessment and/or the child would be distressed through completing the assessment. Detail on exclusion criteria for children will be added to the effectiveness trial parent/carer information sheet and nursery guidance for recruitment.

Additionally, while 33 children receiving EYPP were recruited, only 7 were randomly selected to be assessed. This equates to 8% of the total assessed sample which is lower than the 10% anticipated in protocol (Robinson-Smith et al. 2020). Where nurseries recruit more than the required number of children for the main effectiveness trial, we will consider purposive sampling of children to be assessed using APSECTS, in order that sufficient children eligible for EYPP are included in the sample.

Some nurseries whose intake starts at 3-4 years old were not able to provide data on whether children had previously been in receipt/eligible for FEEE. This data will therefore need to be collected directly from parents at this time of consent in the effectiveness trial. The child details spreadsheet guidance and template will be updated add 'unknown' as a response option (in addition to Y [yes] and N [no]) for FEEE and EYPP eligibility to try to prevent potential reporting errors and reduce missing data.

Unfortunately, we were not able to get feedback directly from parents/carers on potential facilitators and barriers to recruitment, as we had hoped.

Research objective 3: Explore if the intended strategy to reduce attrition is practical, feasible and cost effective.

Our planned strategy to reduce attrition (i.e., assess Cohort 1 children who have left the participating nursery in their new destination nursery or school) was not possible due to the closures of and access to nurseries caused by COVID-19.

In lieu of this data, the evaluation team followed-up nurseries who recruited the younger Cohort 2 sample. Attrition was high; of the 7/9 nurseries who provided data, 22% of children left the nursery. Although the main effectiveness trial will

involve recruiting a different age cohort, strategies to reduce attrition such as assessing children in their 'new destination' may be needed.

The evaluation team will monitor attrition regularly throughout the main effectiveness trial and implement necessary attrition reducing strategies as and when is required.

At the outset of this pilot study, the evaluation team thought SNs may struggle more than PVIs to complete ASPECTS baseline assessments, and therefore offered SNs an RA visit to complete the assessments. Data from the pilot study show that most SN managed to complete baseline assessments themselves. To reduce attrition across the ensure main effectiveness trial, all participating nurseries will be offered an RA visit to complete baseline assessments if they are unable to conduct them in-house.

Research objective 4: Explore the feasibility of recruiting and assessing a 2-3 year old cohort.

The pilot nursery showed that it was possible for nursery staff to 2–3-year-old children; with 40.1% of eligible 2-3 year old recruited to the study.

As part of the exploration of the feasibility of linking to routinely collected ASQ-3 data, we contacted the NHS Digital enquiries team who advised that data linkage may be possible, and requests would need to be made through the Data Access Request Service (DARS). The enquiries team also noted that submitting ASQ-3 data is mandatory and they have recently carried out work to try to encourage higher submission rates of the data.

Through completing the Health Research Authority (HRA) NHS Research Ethics Committee (REC) decision tool, it was determined that NHS REC approval would be required to link to routinely collected ASQ-3 data. As gaining NHS REC approval is a lengthy process (REC require up to 60 days before giving an ethical opinion), to avoid delaying the pilot, it was decided that separate parent/carer information sheets and consent forms would be prepared for ASQ-3 data linkage (in addition to the main information sheets and consent forms), should we choose to go ahead with trying to link to this data within the main effectiveness trial.

One pilot nursery informed the evaluation team that they completed ASQ-3 as standard and uploaded the data to a portal, instead of health visitors completing the ASQ-3 with parents/carers on a home visit. The evaluation team will continue to explore this research objective as part of the main effectiveness trial.

Research objective 5: To explore changes made to the MC programme since the first effectiveness trial and the usefulness and acceptability of these changes within nurseries.

The results of the baseline survey were intriguing, as they revealed that a majority of the nominated MCs possessed qualifications at or above the graduate level. In the previous effectiveness trial, it was mandatory to have a graduate practitioner as the MC (Robinson-Smith et al., 2018). However, for this trial, this requirement was eliminated to align with the evolving landscape of early years practitioner qualifications. . Indeed, NDNA's (2019) annual workforce survey has demonstrated a reduction in the proportion of nursery staff with graduate qualifications in recent years.

It was indicated in the interviews that the Maths Champions programme handbook has been very useful and is beneficial for the nurseries; therefore, it should be sent to the nursery a few days in advance of the online induction.

Findings of the pilot IPE also recommended to establish close contact with a key person, ideally the MC, in each nursery and keep in regular communication with them via email/phone. During the pilot it the delivery team reported that some nurseries were less engaged than others, it was also found that it could be more difficult to maintain contact with SN nurseries, as getting hold of the 'right person' can be more challenging.

Additional Recommendations

Postpone the effectiveness trial by one academic year with programme delivery commencing in September 2021 and continuing until May 2022.

Based on the piloting of the baseline survey, the evaluation team plan to make the following changes to the survey for the main effectiveness trial:

- Add further response options for the question on the role of the respondent (including Nursey owner/director, EYFS Lead, and Senior Nursery Manager), due to the number of respondents who selected 'other' for this question (see Table 3).
- Add further response options for the questions on the role of the nominated MC and DMC (including Senior Nursery Manager, Teacher, Nursery Teacher, EYFS Lead, Higher Level Teaching Assistant), due to the number of respondents who selected 'other' for these questions (see Table 3). It seems that the response options for the survey were more geared to PVI nurseries, rather than maintained or SNs.

It will be useful to explore at the end of the pilot, in the usual practice end point survey, how those nurseries who are running more than one CPD programme managed their capacity to facilitate more than one.

Within the main effectiveness trial, it is recommended that the evaluation team gather information on the possible impact of the pandemic situation on nurseries engagement in the Maths Champions programme.

Recommendations to increase participation in IPE interview within the main effectiveness trial include the delivery team explaining the purpose and importance of the IPE element in the MC programme evaluation during the MC induction session. Additionally, all participant who take part in an IPE interview will receive a £10 high street gift voucher as a thank you for taking part in the interviews.

References

Department of Health and Social Care, 2016. *Factsheet: Developing a public health outcome measure for children aged 2 – 2½ using ASQ-3™*. [online] GOV.UK. Available at: <https://www.gov.uk/government/publications/measuring-child-development-at-age-2-to-25-years> [Accessed 16 February 2023].

Early Education, 2015. *Maintained nursery schools: the state of play report*. [online] London: Early Education. Available at: <https://www.early-education.org.uk/sites/default/files/Nursery%20Schools%20State%20of%20Play%20Report%20final%20print.pdf> [Accessed 8 July 2020].

Early Education, 2018. *Briefing on funding for maintained nursery schools*. [online] Available at: <https://www.early-education.org.uk/sites/default/files/Briefing%20on%20funding%20for%20Maintained%20Nursery%20Schools%20June%202018.pdf> [Accessed 8 July 2020].

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K., 2016. *Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook*. [online] London: Education Endowment Foundation. Available at: <https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Handbook.pdf>.

National Education Union, 2020. *Save maintained nursery schools*. [online] Available at: <https://neu.org.uk/funding/save-maintained-nursery-schools> [Accessed 8 July 2020].

NDNA, 2019. *NDNA 2018/19 Workforce Survey England*. [online] Huddersfield: National Day Nurseries Association. Available at: <https://www.ndna.org.uk/NDNA/News/Reports_and_surveys/Workforce_survey/nursery_workforce_survey_2019.aspx> [Accessed 20 November 2019].

Paull, G. and Popov, D., 2019. *The role and contribution of maintained nursery schools in the early years sector in England: Research report*. [online] London: Department for Education. Available at: <https://www.gov.uk/government/publications/maintained-nursery-schools-contribution-to-early-years-provision> [Accessed 8 July 2020].

Public Health England, 2018. *Feasibility study: developing the capability for population surveillance using indicators of child development outcomes aged 2 to 2 and a half years*. [online] London: Public Health England. Available at: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/683601/Feasibility_study_developing_the_capability_for_population_surveillance_using_indicators_of_child_development_outcomes_aged_2_to_2_and_a_half_years.pdf> [Accessed 12 July 2020].

Public Health England, 2020. *Child development outcomes at 2 - 2½ years (Experimental Statistics): Quarter 3 2019/20 Statistical Commentary (April 2020 release)*. [online] London: Public Health England. Available at: <https://www.gov.uk/government/statistics/child-development-outcomes-at-2-to-2-and-a-half-years-2019-to-2020> [Accessed 12 July 2020].

Robinson-Smith, L., Fairhurst, C., Stone, G., Bell, K., Elliott, L., Gascoine, L., Hallett, S., Hewitt, C., Hugill, J., Torgerson, C., Torgerson, D., Menzies, V. and Ainsworth, H., 2018. *Maths Champions: Evaluation report and executive summary*. [online] London: Education Endowment Foundation. Available at: <https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Maths_champions_evaluation_report.pdf> [Accessed 8 October 2019].

Robinson-Smith, L., Whiteside, K., Torgerson, C., Fairhurst, C., Podmore, D., Davill, T., Bell, K., Qi, X., Leggett, H., Elliott, L., Hewitt, C., Baird, K., Menzies, V., Torgerson, D. and Ainsworth, H., 2023. *Independent evaluation of Maths Champions in nursery to develop children's early numeracy: A two-armed cluster randomised controlled trial. Evaluation Report [in preparation]*. London: Education Endowment Foundation.

Robinson-Smith, L., Whiteside, K., Torgerson, C., Qi, X., Fairhurst, C., Elliott, L., Hewitt, C., Joshi, K., Menzies, V., Torgerson, D. and Ainsworth, H., 2022. *Independent evaluation of Maths Champions in nursery settings to develop*

children's early numeracy: A two-armed cluster randomised controlled trial. Evaluation Protocol. [online] London: Education Endowment Foundation. Available at: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/maths-champions-effectiveness/> [Accessed 12 December 2022].

Squires, J. and Bricker, D., 2009. *Ages and Stages Questionnaire, Third Edition (ASQ-3)*. Paul H. Brookes Publishing Co.

Appendix F – Long-term NPD EYFSP outcomes

Introduction

This is an appendix to the evaluation report for 'Independent evaluation of Maths Champions in nursery to develop children's early numeracy: A two-armed cluster randomised controlled trial'. The original trial related to 1,304 children from 134 nurseries aged three to four years in nursery classes during the 2021/2022 academic year. The trial randomised nurseries to either receive the Maths Champions (MC) programme (intervention group) or to continue with usual nursery provision (control group). The impact of MC on children's mathematics attainment was evaluated.

This appendix reports on the long-term outcomes collected from the National Pupil Database (NPD) for children recruited into the MCII trial. Early Years Foundation Stage Profile (EYFSP) assessment data for Mathematics, Literacy, and 'Good Level of Development' (GLD) were used to measure impact. The EYFSP is a teacher-reported assessment undertaken at the end of Reception—the first year of statutory education. This assessment was completed in the summer term of the 2022/2023 academic year. Data were requested from the NPD for all 1,304 participating children.

Evaluation objectives

The longitudinal analysis research questions (LRQs) were:

LRQ 1. What is the impact of the MC programme, in comparison to usual early years setting provision, on the mathematical development of children at the end of Reception, as measured by the two mathematical early learning goals of the EYFSP?

LRQ 2. What is the impact of the MC programme, in comparison to usual early years setting provision, on the literacy of children at the end of Reception, as measured by the three literacy early learning goals of the EYFSP?

LRQ 3. What is the impact of the MC programme, in comparison to usual early years setting provision, on children's overall development and school readiness, as measured by whether the child achieved a Good Level of Development score in the EYFSP?

Appendix F Table 32. Summary of impact on primary and longitudinal outcomes

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils (intervention; control)	P value	EEF cost rating
Primary outcome: Mathematics (ASPECTS)	0.25 (0.12 to 0.38)	3	🔒🔒🔒🔒🔒	1,209 (600; 609)	0.001	**£** £ £ £ £
Longitudinal outcomes						
EYFSP Mathematics[a]	0.13 (-0.11, 0.36)	2	N/A	1,271 (624; 647)	0.31	N/A
EYFSP Literacy[b]	0.10 (-0.14, 0.34)	2	N/A	1,271 (624; 647)	0.39	N/A
EYFSP Good Level of Development[a]	0.08 (-0.16, 0.32)	1	N/A	1,277 (624; 653)	0.50	N/A
EYFSP Good Level of Development[b]	0.09 (-0.14, 0.31)	1	N/A	1,277 (624; 653)	0.44	N/A
Subgroup analysis						

EYPP subgroup: EYFSP Mathematics[a]	0.45 (-0.06, 0.95)	6	N/A	165 (69; 96)	0.08	N/A

[a] adjusted for baseline mathematics ASPECTS

[b] adjusted for baseline language ASPECTS

Appendix F Table 33. Key conclusions

Key conclusions	
1.	Children in nurseries allocated to the intervention group made, on average, the equivalent of two months' additional progress in mathematics attainment compared to children in control nurseries as measured by the EYFSP.
2.	Children in nurseries allocated to the intervention group made, on average, the equivalent of two months' additional progress in literacy attainment compared to children in control nurseries as measured by the EYFSP.
3.	Children in nurseries allocated to the intervention group were slightly more likely to attain a Good Level of Development (GLD) compared to children in control nurseries as measured by the EYFSP (equating to approximately one months' additional progress).
4.	Children eligible for Early Years Pupil Premium (EYPP) in the intervention nurseries made, on average, the equivalent of six months' additional progress in mathematics attainment compared to children eligible for EYPP in control nurseries. These results, while promising, should be treated with more caution than the analysis on all pupils as fewer children were included in this analysis.

Methods

Addendum outcome measures

The EYFSP is an observational measure completed by teachers. Teachers rate each child's learning and development against 17 early learning goals (ELGs) using the following two levels: meeting the level of development expected at the end of the EYFS (expected); or not yet reaching this level (emerging). For any of the ELGs, a score of 'A' may be reported to indicate that a child has not been assessed.

Mathematics is a specific area of learning measured in the EYFSP, using two ELGs: Number; and Numerical Patterns. A binary measure of whether or not the pupil met 'expected' levels in both of these ELGs was an outcome for the longitudinal analysis and relates to LRQ 1.

Literacy is another learning area measured by the EYFSP, using three ELGs: Comprehension; Word Reading; and Writing. A binary measure of whether or not the pupil met 'expected' levels in all three of these ELGs was an outcome for the longitudinal analysis and relates to LRQ 2.

'Good Level of Development' (GLD) is a dichotomous variable (Yes/No) pre-calculated and provided as a single variable in the NPD. Children are defined as having reached a GLD at the end of the EYFS if they have achieved at least the expected level for the ELGs in the prime areas of learning (communication and language; physical development; and personal, social and emotional development) and the specific areas of mathematics and literacy identified above. This outcome for the longitudinal analysis relates to LRQ 3.

In the MCII study, the intervention period ended in July 2022 (and took place in nurseries when the children were 3-4 years old) and the EYFSP was conducted one year later in schools, as usual, in July 2023 (taking place at the end of Reception when children were 4-5 years old). As a result, the EYFSP was conducted by teachers in schools, who had no role in delivering the intervention, which was delivered by nursery staff. The results of the EYFSP are made available within the NPD in November, following the assessments in the summer, for researchers to request access to the data. The research team accessed the results for children participating in the MCII evaluation to assess the longer-term impact of the intervention. An application for NPD data was submitted to the Department for Education (DfE) via the Office for National Statistics (ONS) Secure Research Service (SRS) in September 2023. The research team provided pupil details—pupil trial ID, forename, surname, date of birth, home postcode, nursery trial id, nursery Unique Reference

Number (URN) and local authority establishment number (LAESTAB), intended primary school destination name, post code, and URN (children are only assigned a unique pupil number (UPN) when they reach primary school, so this was not available to match on)—for the DfE to match to the requested NPD data. The export of NPD data was transferred to the ONS SRS for the evaluation team at University of York to analyse and did not contain personal data. The only (meaningless) identifiers transferred were the pupil matching reference (PMR) and the pupil and nursery trial IDs. Data from the main evaluation was then transferred to the SRS early in 2024 so it could be merged, via the pupil trial ID, with the NPD data for analysis.

The trial statistician, Caroline Fairhurst, is an ONS accredited researcher; she submitted the application and accessed and analysed the data.

Analysis

Analyses were conducted on an intention-to-treat basis (ITT), using two-sided significance at the 5% level, using Stata v17. The published statistical analysis plan can be found here, and follows the latest EEF analysis guidance. Nursery and child-level characteristics and baseline data are summarised descriptively by randomised group for participants for whom EYFSP data is available. Outcome data are summarised descriptively for the two groups. The correlations between EYFSP measures and ASPECTS baseline measures (collected during the main trial) are reported with a 95% confidence interval (CI). The research team originally hoped to collect ASQ-3 data in the main evaluation and calculate the correlation between this and EYFSP measures; however, as detailed in the main report, collection of ASQ-3 data was ultimately determined to be infeasible.

Longitudinal analyses

The three dichotomous outcomes were analysed via mixed-effect logistic regression, adjusted for baseline ASPECTS score (numeracy for mathematics outcome, language for literacy outcome, and both (separately) for the GLD outcome), and setting-level minimisation factors (nursery type—PVI or SN/maintained, nursery size, and number of staff at the nursery holding a degree qualification in early years) as fixed effects, and nursery as a random effect. The continuous variables that were dichotomised to use as factors in the minimisation procedure (nursery size and number of graduate nursery staff) were included in their continuous form in the models. The GLD analysis model was run twice, adjusting for both the mathematics and language ASPECTS score separately as a measure of prior attainment. This is because the GLD incorporates measures of both numeracy and literacy and so we were unsure which score would correlate more highly with the GLD outcome, and thus potentially produce a more precise result.

The raw percentage point difference between the proportion of participants in each group achieving the outcome is presented with a 95% CI. From the mixed-effect logistic model, the treatment effect expressed as an odds ratio (OR; adjusted for relevant baseline ASPECTS and setting-level minimisation factors) is reported with a 95% CI and p-value. The adjusted OR (and 95% CI limits) is converted to an approximation of Hedges' g effect size (to facilitate interpretation and provide comparability with the main trial results, which were based on continuous outcomes) via the Cox index as follows (What Works Clearinghouse):

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

Where $\omega = \left[1 - \frac{3}{4N-9}\right]$ and *N* is the total sample size. The approximated Hedges' g is, in turn, related to an estimated 'months of progress', as per the EEF evaluation report template. As a small note of caution, these results should therefore be considered in the context that they have been obtained via a series of approximations.

The predicted percentage point difference between the two trial arms (estimated from the logistic regression model using the postestimation Stata command *margins, dydx(allocation)*) is also presented with a 95% CI (Ge et al., 2011). The presentation of the treatment effect estimate as a percentage point difference and converted Hedges' g effect size is recommended by the EEF analysis guidance to complement the OR and aid interpretation.

The intra-cluster correlation coefficient (ICC) associated with nursery is also presented for each outcome alongside a 95% CI. Both the conditional ICC, obtained from the adjusted analysis model, and the unconditional ICC, obtained from an empty model (i.e. one without covariates), are presented.

Subgroup analysis

In line with the effectiveness trial analysis, subgroup analyses as part of the longitudinal evaluation considered children that were eligible for the EYPP, FEEE at two years old, average number of hours that the child attends nursery, gender and PVI vs school-based nurseries. These analyses were only undertaken for the outcome for the longitudinal analysis of mathematics attainment. The subgroup analyses were conducted by including the factor and an interaction term between the factor and allocation in the analysis model. The longitudinal analysis for the mathematics outcome was also repeated within the subset of participants eligible for EYPP.

Impact evaluation

Attrition

NPD data were requested for all 1,304 children (intervention, n = 638; control, n = 666) from 134 nurseries (intervention, n = 66; control, n = 68), of which a record for the child in the NPD could be identified (based on the personal information provided by the evaluation team e.g. name, date of birth, home postcode, etc) for 1,277. In total, 1,271 (97.5%; intervention, n = 624, 97.8%; control, n = 647, 97.1%) children were included in the longitudinal analysis for the mathematics outcome (from all 134 randomised nurseries) as they had valid EYFSP data for all ELGs (no "A" ratings). Therefore, the overall attrition rate was 2.5% (intervention group: 2.2%; control group: 2.9%). This is a lower level of attrition than was observed in the primary analysis in the main evaluation in which 1,209 pupils were included, so the final overall attrition rate was 7.3% (intervention group: 6.0%; control group: 8.6%). The median number of children included in the longitudinal analysis model for the mathematics outcome from each of the 134 nurseries was 10 in both the intervention and control groups.

Pupil characteristics of analysed sample

The characteristics of the 134 nurseries and 1,271 children in the analysed sample are presented in Appendix F Table 34 and are broadly similar between the intervention and the control groups. Since no formal hypothesis testing was performed on baseline data, comparisons are made by eye only.

Nurseries in the intervention group had, on average, 2.2 graduate staff members; 61 (92.4%) nurseries had at least one graduate staff member. Similarly, the control group had, on average, 2.4 graduate staff members per nursery, with 63 (92.6%) having at least one graduate staff member.

There were 35 (53.0%) PVI nurseries in the intervention group and 36 (52.9%) in the control group, and the rest were maintained or SN. The majority were private or independent—32 (48.5%) in the intervention group and 32 (47.1%) in the control group—and a small number were voluntary: three (4.5%) in the intervention group and four (5.9%) in the control group.

Approximately half of the children were male: 306 (49.0%) in the intervention group and 320 (49.5%) in the control group. Regarding EYPP eligibility status among children, there were 49 (7.9%) missing datapoints in the intervention group and 43 (6.7%) in the control group. For children with EYPP data, 69 (12.0%) in the intervention group were eligible for EYPP and 96 (15.9%) in the control group were eligible for EYPP.

Some differences were observed for English as an Additional Language (EAL) between groups. There is a larger proportion of missing data in the control group compared to the intervention group: 69 (10.7%) versus 30 (4.8%). For children with EAL data, the control group had a larger proportion of EAL children; in the intervention group there were 51 (8.6%) children with EAL and in the control group there was 78 (13.5%). There is also a larger proportion of children eligible for FEEE at two years old (defined by the composite FEEE variable) in the control group (38.5%) than in the intervention group (31.9%).

The baseline table also reports on the EYFS two-year check (conducted when children are between 24-36 months old), which supplies parents and carers with a short, written summary of their child's development in the three prime learning and development areas of the EYFS: Personal, Social and Emotional Development; Physical Development; and Communication and Language. It provides parents and carers with an estimate of their child's developmental age at the time of the assessment. These assessments took place before the ASPECTS baseline assessment in the main trial and were completely independent of the evaluation. Data were provided by nurseries but were not provided by every nursery or for every child within a nursery (depending on whether the child was attending the nursery at the time of the

assessment); hence, there is a large amount of data for these variables. They should not be confused with the EYFSP outcomes being used in this longitudinal analysis, which were assessed when the children were aged four to five years old. Where data were provided, there appears to be a higher proportion of children in the older age bracket of 30-60 months development in the control group than the intervention group across all three learning and development areas.

The characteristics of the pupils analysed in this longitudinal follow-up were virtually identical to those randomised and those analysed in the primary analysis for the main evaluation (see Tables 10 and 11 of main report).

Appendix F Table 34. Baseline nursery and child characteristics for the population as analysed in the longitudinal evaluation

Nursery level (categorical)	**Intervention group**		**Control group**	
	n/N (missing)	**Count (%)**	**n/N (missing)**	**Count (%)**
>1 graduate staff	66/66 (0)	61 (92.4)	68/68 (0)	63 (92.6)
Nursery type PVI SN and Maintained	66/66 (0)	 35 (53.0) 31 (47.0)	68/68 (0)	 36 (52.9) 32 (47.1)
Nursery type subgroups Private/independent School-based Maintained Voluntary	66/66 (0)	 32 (48.5) 21 (31.8) 10 (15.2) 3 (4.5)	68/68 (0)	 32 (47.1) 20 (29.4) 12 (17.6) 4 (5.9)
Nursery level (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**
Number of graduate staff	66/66 (0)	2.2 (1.6)	68/68 (0)	2.4 (1.9)
Number of eligible children	66/66 (0)	35.4 (22.4)	68/68 (0)	33.5 (16.6)
Child-level (categorical)	**n/N (missing)**	**Count (%)**	**n/N (missing)**	**Count (%)**
Sex, male	624/624 (0)	306 (49.0)	647/647 (0)	320 (49.5)
Eligible for EYPP	575/624 (49)	69 (12.0)	604/647 (43)	96 (15.9)
EAL	594/624 (30)	51 (8.6)	578/647 (69)	78 (13.5)
Eligible for FEEE at 2 years old (parent/carer response)	559/624 (65)	162 (29.0)	576/647 (71)	203 (35.2)
Received FEEE at 2 years old (parent/carer response)	615/624 (9)	181 (29.0)	642/647 (5)	230 (35.5)
Received FEEE at 2 years old (nursery response)	563/624 (61)	116 (20.6)	537/647 (110)	138 (25.7)
Eligible for FEEE at 2 years old (composite of above 3 variables)	624/624 (0)	199 (31.9)	647/647 (0)	249 (38.5)

EYFS2Y personal, social, emotional development [a] 0–20m 16–26m 22–36m 30–60m	293/624 (331)	 12 (4.1) 58 (19.8) 185 (63.1) 38 (13.0)	266/647 (381)	 11 (4.1) 54 (20.3) 160 (60.2) 41 (15.4)	
EYFS2Y Communication and Language [a] 0–20m 16–26m 22–36m 30–60m	295/624 (329)	 15 (5.1) 66 (22.4) 173 (58.6) 41 (13.9)	267/647 (380)	 11 (4.1) 62 (23.2) 146 (54.7) 48 (18.0)	
EYFS2Y Physical development [a, b] 0–26m 22–36m 30–60m	293/624 (331)	 67 (22.9) 183 (62.5) 43 (14.7)	267/647 (380)	 49 (18.4) 162 (60.7) 56 (20.2)	
Child level (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**Hedges' *g* effect size (95% CI)**
Hours attendance at nursery per week	624/624 (0)	24.5 (8.7)	647/647 (0)	24.3 (9.3)	N/A
Baseline ASPECTS mathematics score	624/624 (0)	12.6 (6.9)	647/647 (0)	12.8 (6.6)	0.04 (-0.07, 0.15)
Baseline ASPECTS language score	624/624 (0)	24.4 (9.0)	647/647 (0)	24.5 (8.3)	0.02 (-0.08, 0.13)

[a] The age ranges (in months) refer to developmental age of the child at the time the EYFS2Y check, as reported by the completing practitioner.
[b] Cell counts less than ten suppressed to prevent statistical disclosure.

Analysis

Summary of raw scores

Outcome scores are summarised in

Appendix F Table 35. A total of 1,271 children had a valid rating for both the Number and Numerical Patterns ELG so a composite mathematics outcome could be calculated for them, of which a total of 1,115 (87.7%) were rated as 'Expected' on both (intervention, n = 551, 88.3%; control, n = 564, 87.2%).

A total of 1,271 children had a valid rating for the Comprehension, Word reading, and Writing ELGs so a composite literacy outcome could be calculated for them, of which, 1,030 (81.0%) children were rated 'Expected' on all three ELGs (intervention, n = 510, 81.7%; control, n = 520, 80.4%).

A GLD outcome (pre-defined in the NPD) was available for 1,277 pupils (children with "A" rated ELGs were given a rating of not having met the threshold for GLD). A slightly higher proportion of children achieved a GLD in the EYFSP in the intervention group (n = 497, 79.6%) than the control group (n = 510, 78.1%).

Appendix F Table 35. Summary of raw outcome scores from EYFSP for longitudinal analysis

Variable	Intervention group (n = 624)	Control group (n = 647)	Overall (n = 1,271)
Mathematics area of learning from EYFSP, n (%)			
Number ELG			
Emerging	61 (9.8)	73 (11.3)	134 (10.5)
Expected	563 (90.2)	574 (88.7)	1137 (89.5)
Numerical Patterns ELG			
Emerging	71 (11.4)	78 (12.1)	149 (11.7)
Expected	553 (88.6)	569 (87.9)	1122 (88.3)
Met expected level in both	551 (88.3)	564 (87.2)	1115 (87.7)
Literacy area of learning from EYFSP, n (%)			
Comprehension ELG			
Emerging	50 (8.0)	58 (9.0)	108 (8.5)
Expected	574 (92.0)	589 (91.0)	1163 (91.5)
Word Reading ELG			
Emerging	82 (13.1)	102 (15.8)	184 (14.5)
Expected	542 (86.9)	545 (84.2)	1087 (85.5)
Writing ELG			
Emerging	109 (17.5)	122 (18.9)	231 (18.2)
Expected	515 (82.5)	525 (81.1)	1040 (81.8)
Met expected level in all three	510 (81.7)	520 (80.4)	1030 (81.0)
	Intervention group (n = 624)	**Control group (n = 653)**	**Overall (n = 1,277)**
Achieving a good level of development, n (%)			
Yes	497 (79.6)	510 (78.1)	1007 (78.9)

Longitudinal analysis

Mathematics

The correlation between the baseline ASPECTs mathematics score and EYFSP mathematics outcome was positive and moderate in size; calculated as 0.37 (95% CI 0.32 to 0.42). The difference between the intervention and control groups for the mathematics outcome favoured the intervention group; the raw percentage point difference was 1.13 (95% CI -2.48 to 4.74). The estimated Hedges' g effect size was 0.13 (95% CI -0.11 to 0.36), equating to two months'

additional progress in the intervention group. The 95% CI suggests the true effect could range from two months' fewer progress to five months' additional progress in the intervention group and includes 0 so the difference was not statistically significant (p = 0.31) (Appendix F Table 36). The ICC associated with nursery setting from the adjusted model was 0.03 (95% CI 0.001 to 0.44). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.09 (0.03 to 0.22).

Literacy

The correlation between the baseline ASPECTs language score and EYFSP literacy outcome was positive and moderate; calculated as 0.36 (95% CI 0.31 to 0.40). The difference between the intervention and control groups for the literacy outcome favoured the intervention group; the raw percentage point difference was 1.36 (95% CI -2.95 to 6.78). The estimated Hedges' g effect size was 0.10 (95% CI -0.14 to 0.34), equating to two months' additional progress in the intervention group. The 95% CI suggests the true effect could range from two months' fewer progress to four months' additional progress in the intervention group and includes 0 so the difference was not statistically significant (p = 0.39) (Appendix F Table 36). The ICC associated with nursery setting from the adjusted model was 0.10 (95% CI 0.04 to 0.23). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.10 (0.05 to 0.20).

EYFSP Good Level of Development

Children in the intervention group were slightly more likely to have achieved a GLD than those in the control group; the raw percentage point difference was 1.55 (95% CI -2.93 to 6.02). The correlation between the GLD outcome and the two baseline ASPECTS scores (mathematics and language) were similar, and both indicated a moderate, positive association. In the model adjusting for baseline ASPECTS mathematics score, which correlated marginally more highly with this outcome than baseline ASPECTS language score, the estimated Hedges' g effect size was 0.08 (95% CI -0.16 to 0.32) or one months' additional progress (Appendix F Table 36). The 95% CI suggests the true effect could range from two months' fewer progress to four months' additional progress in the intervention group and includes 0 so the difference was not statistically significant (p = 0.50). This was similarly true when the GLD analysis model was adjusted for baseline ASPECTS language score. The ICC associated with nursery setting was 0.14 (95% CI 0.07 to 0.25) and 0.11 (95% CI 0.05 to 0.22) for the models adjusted for baseline ASPECTS mathematics and language, respectively. The ICC (95% CI) for the empty model (i.e. without covariates) was 0.12 (0.06 to 0.21).

Appendix F Table 36. EYFSP longitudinal outcomes for Maths Champions II (n=1271)

	Correlation with baseline ASPECTS (95% CI)	Unadjusted percentage point difference (95% CI)	Adjusted percentage point difference (95% CI)	Adjusted OR (95% CI)	Hedges' g (95% CI)	p-value
EYFSP Mathematics[a]	0.37 (0.32, 0.42)	1.13 (-2.48, 4.74)	1.77 (-1.66, 5.19)	1.23 (0.83, 1.82)	0.13 (-0.11, 0.36)	0.31
EYFSP Literacy[b]	0.36 (0.31, 0.40)	1.36 (-2.95, 5.67)	2.06 (-2.65, 6.78)	1.18 (0.80, 1.74)	0.10 (-0.14, 0.34)	0.39
EYFSP GLD[a]	0.37 (0.32, 0.42)	1.55 (-2.93, 6.02)	1.77 (-3.42, 6.96)	1.15 (0.77, 1.71)	0.08 (-0.16, 0.32)	0.50
EYFSP GLD[b]	0.34 (0.29, 0.39)	1.55 (-2.93, 6.02)	1.98 (-3.05, 7.02)	1.16 (0.80, 1.68)	0.09 (-0.14, 0.31)	0.44

[a] adjusted for baseline mathematics ASPECTS

[b] adjusted for baseline language ASPECTS

Subgroup analysis

Summary statistics for the EYFSP mathematics outcome are presented in Appendix F Table 37 by gender, average number of hours the child attends the nursery per week (dichotomized at the median of 24 hours), eligibility for EYPP, eligibility for FEEE at two years old, and nursery type. These are displayed as the number and percentage of children within each level of the subgroup that met 'expected' levels for both mathematics ELGs. These summaries indicate

that, in general, children attending nursery for more than 24 hours per week and children attending PVI nurseries were more likely to meet this level than those who attended nursery less than 24 hours a week or who attended a school-based nursery, respectively. Children eligible for EYPP or FEEE were less likely to meet this level than children who were not eligible.

In a series of adjusted regression analyses that included interaction effects, there was no evidence of an interaction between trial allocation and (i) gender (interaction term, coefficient 1.23, 95% CI 0.57 to 2.62, p = 0.60), (ii) average number of hours child attends nursery per week (dichotomised at 24 hours, 0.74, 95% CI 0.34 to 1.61, p = 0.44), (iii) eligibility for EYPP (0.61, 95% CI 0.22 to 1.66, p = 0.34), (iv) eligibility for FEEE (0.61, 95% CI 0.28 to 1.32, p = 0.21), or (v) type of nursery child attended (PVI vs SN; 0.67, 95% CI 0.30 to 1.46, p = 0.31).

There were 165 children from 63 nurseries who were eligible for EYPP and had a valid baseline ASPECTS mathematics score and EYFSP mathematics outcome (median two per nursery, range one to eight in both groups). Within the restricted sample of children eligible for EYPP, the estimated Hedges' g effect size was 0.45 (95% CI -0.06 to 0.95, p = 0.08), equating to six months' additional progress in the intervention group (Appendix F Table 38); however, the 95% CI suggests the true effect could lie in the range from one month's fewer progress to 11 months' additional progress in the intervention group. These results are based on a small number of participants, hence the wide confidence intervals, and so should be interpreted as exploratory. The ICC associated with nursery setting from the adjusted model could not be calculated due to the small number of children per cluster.

Appendix F Table 37. Subgroup summary scores for the EYFSP mathematics outcome

	Intervention	**Control**
Outcome	**N (%)**	**N (%)**
Gender Male Female	270 (88.2) 281 (88.4)	282 (88.1) 282 (86.2)
Number of hours at nursery < 24 ≥ 24	252 (86.3) 299 (90.1)	271 (83.1) 293 (91.3)
Eligible for EYPP Yes No	55 (79.7) 458 (90.5)	65 (67.7) 462 (90.9)
Eligible for FEEE at 2 years old Yes No	160 (80.4) 391 (92.0)	197 (79.1) 367 (92.2)
Nursery type PVI SN	299 (90.3) 252 (86.0)	301 (87.5) 263 (86.8)

Appendix F Table 38. Result for EYPP subgroup analysis for the mathematics outcome (n=165)

	Correlation with baseline ASPECTS (95% CI)	**Unadjusted percentage point difference (95% CI)**	**Adjusted percentage point difference (95% CI)**	**Adjusted OR (95% CI)**	**Hedges' g (95% CI)**	**p-value**
EYFSP Mathematics	0.49 (0.36, 0.60)	12.00 (-1.32, 25.33)	11.28 (-0.84, 23.40)	2.10 (0.91, 4.83)	0.45 (-0.06, 0.95)	0.08

Conclusion

Appendix F Table 39. Key conclusions

Key conclusions	
1.	Children in nurseries allocated to the intervention group made, on average, the equivalent of two months' additional progress in mathematics attainment compared to children in control nurseries as measured by the EYFSP.
2.	Children in nurseries allocated to the intervention group made, on average, the equivalent of two months' additional progress in literacy attainment compared to children in control nurseries as measured by the EYFSP.
3.	Children in nurseries allocated to the intervention group were slightly more likely to attain a Good Level of Development (GLD) compared to children in control nurseries as measured by the EYFSP (equating to approximately one months' additional progress).
4.	Children eligible for Early Years Pupil Premium (EYPP) in the intervention nurseries made, on average, the equivalent of six months' additional progress in mathematics attainment compared to children eligible for EYPP in control nurseries. These results, while promising, should be treated with more caution than the analysis on all pupils as fewer children were included in this analysis.

Interpretation

The MC intervention aimed to improve the mathematical and language attainment of children aged three to four years old and the main evaluation observed increases equating to approximately three months' additional progress (95% CI two to five months). MCII was a rigorously designed, conducted, and reported cluster randomised controlled trial. Results from the main evaluation were considered to have a very high security rating. This pre-planned longitudinal analysis found that participation in the MC intervention also led to small improvements in mathematical and language skills at age four to five years old; children attending nurseries allocated to receive the MC programme made an estimated two months' additional progress in mathematics and literacy as measured by the EYFSP at the end of their first year of primary school, although both results could not rule out the possibility that the true effect favours the control group as the 95% CI includes the range from two months' fewer progress to five or four months' additional progress (respectively for mathematics and literacy) in the intervention group.

Children in the MC group were slightly more likely to achieve a 'good level of development', equating to approximately one month's additional progress, though again the 95% CI could not rule out a negative result (two months' fewer progress to four months' additional progress in the intervention group).

Within the longitudinal subgroup analysis, children in the intervention group who were eligible for EYPP were observed to make six months' additional progress in mathematics attainment compared to children in the control group who were eligible for EYPP. This result was maintained from the main evaluation.

Children from all 134 randomised nurseries were included in this longitudinal analysis and, in terms of child-level attrition, we were able to obtain data for a higher proportion of children in this analysis than the primary analysis in the main evaluation. At the level of the child, attrition was very low and similar between the two groups in the main evaluation - 6.0% and 8.6% for the intervention and control groups, respectively - and was even lower here (2.5% overall, intervention group 2.2%; control group 2.9%). Nursery and child-level characteristics for the 'as analysed' sample were very similar to the 'as randomised' sample. Therefore, there was no evidence that attrition introduced selection bias.

In summary, this longitudinal analysis found small, positive effects of the intervention in all outcomes, particularly among children eligible for EYPP. The findings would suggest that some of the benefits of the MC intervention are maintained after children have left nursery and entered school.

Limitations

The EYFSP is teacher-assessed. This meant that the EYFSP was assessed by someone not involved in the delivery of the intervention. However, it is possible that some teachers were aware of the trial arm the child was in. For example, for children who previously attended a school-based nursery and subsequently attended the associated primary school,

the Reception teacher may have been aware of whether the nursery had been randomised to receive the MC programme or not, and therefore whether the child had been exposed to the intervention. We are unable to obtain data to quantify this, but we believe the chance of this to influence the results to be very low. Even if teachers were aware of trial arm, it is unlikely they were also aware that we were planning to use EYFSP data as a longitudinal outcome so the risk of bias in assessments is negligible.

The primary outcome result for ASPECTS mathematics score in the main evaluation was a predicted benefit of approximately three months' additional benefit, with a 95% CI that did not include zero. The result for this longitudinal analysis of EYFSP mathematics outcome was of a slightly lower magnitude (two months) with wider confidence limits (minus two months to five months). A smaller effect was perhaps to be expected due to the increased time since intervention receipt. The decrease in precision may be due to the lower correlation between baseline ASPECTS mathematics score and the EYFSP outcome (of 0.37) than between baseline and outcome ASPECTS mathematics score (of 0.61). Again, this may be due to the fact that more time has elapsed between the measurement of these assessments.

Ultimately, we estimate the MDES for the longitudinal analysis of the mathematics outcome to be approximately 0.16, based on a correlation of 0.37, ICC of 0.03 and average cluster size of 9.5 children per nursery. Therefore, the analysis was underpowered to detect an effect size of 0.13.

The outcomes used here were binary and so the treatment effect could not be presented as a Hedges' g effect size. However, to facilitate interpretation and provide comparability with the main trial results, which were based on continuous outcomes, a transformation from odds ratio to the Cox Index, which is comparable to Hedges' g, has been used. This is turn has been related to an estimated 'months of progress'. Results should therefore be interpreted in the context that they have been obtained through a series of approximations.

Findings from the EYPP subgroup analysis are based on a relatively small number of children (approximately 13% of the full sample) and so, while promising, should be interpreted with caution.

Future research

Future research could continue to track the educational progress of children in the MCII evaluation through other standard national tests, e.g. Key Stage 1 and 2 assessments, and GCSEs. This could provide evidence on whether, and for how long, improvements in mathematics and language/literacy are maintained for children in the intervention group.

ONS Disclaimer

References

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

What Works Clearinghouse (n.d). Procedures Handbook, Version 4.0, p.13-14: https://ies.ed.gov/ncee/wwc/docs/referenceresources/wwc_procedures_handbook_v4.pdf

Further appendices

Please see accompanying document *'Technical Notes'*.

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

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

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