Education Endowment Foundation

Maths Champions

Evaluation report and executive summary

July 2018

Independent evaluators:

Lyn Robinson-Smith, Caroline Fairhurst, Gemma Stone, Kerry Bell, Louise Elliott, Louise Gascoine, Sarah Hallett, Catherine Hewitt, Jessica Hugill, Carole Torgerson, David Torgerson, Victoria Menzies, Hannah Ainsworth

University of York

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.
The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education. Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

This project was jointly funded by the EEF and the Stone Family Foundation.

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

Danielle Mason
Head of Research

Education Endowment Foundation, 9th Floor, Millbank Tower, 21–24 Millbank, SW1P 4QP
p: 020 7802 1679
e: danielle.mason@eefoundation.org.uk
w: **www.educationendowmentfoundation.org.uk**

About the evaluator

The project was independently evaluated by a team from York Trials Unit, University of York and Durham University: Lyn Robinson-Smith, Caroline Fairhurst, Gemma Stone, Kerry Bell, Louise Elliott, Louise Gascoine, Sarah Hallett, Catherine Hewitt, Jessica Hugill, Carole Torgerson, David Torgerson, Victoria Menzies, and Hannah Ainsworth.

The lead evaluators were Hannah Ainsworth and Victoria Menzies.

Contact details:

Name: Hannah Ainsworth
Address: York Trials Unit, ARRC Building, University of York, York, YO10 5DD
Tel: 01904 328158
Email: hannah.ainsworth@york.ac.uk

Contents

Executive summary

The project

Maths Champions aims to improve the maths skills of children in private, voluntary and independent (PVI) early years settings. It is a one-year programme developed and delivered by the National Day Nurseries Association (NDNA), supported in this project by Sandra Mathers and Dr Maria Evangelou at the University of Oxford.

Each setting nominates a graduate practitioner to be a "Maths Champion". NDNA provides each Maths Champion with two, two-hour online courses on auditing early years maths teaching and leading the programme in their setting. The Maths Champions then audit their nursery's current practices, staff confidence, and the current mathematical competence of both staff and children using tools provided online. The Maths Champions are expected to use the results of the audit to create an action plan for supporting their colleagues to improve their maths teaching. They implement this plan with support from online resources, short monthly webinars and tailored one-to-one support provided by NDNA. At the end of the programme, the Math Champions repeat the initial audit to gauge their setting's improvement.

Maths Champions was evaluated using a randomised controlled trial (RCT). 108 nurseries were randomly allocated to either participate in Maths Champions or continue with their "business as usual" provision. The primary outcome was performance on a maths test of children who were three years old at the start of the programme. A process evaluation used case studies, surveys, interviews with the project delivery team, observation, and consideration of monitoring data collected by NDNA. The evaluation took place between May 2016 and August 2017, with most nursery settings allocated to the intervention group completing the audits between August and October 2016 and delivering Maths Champions between September/October 2016 and June/July 2017. It was preceded by a small-scale pilot, which involved 6 PVI nurseries and lasted from January 2016 until April 2016.

This project was jointly funded by the EEF and the Stone Family Foundation.

Key conclusions
1. Children who participated in Maths Champions made the equivalent of two additional months' progress in maths, on average, compared to children in the comparison group. This result has low security and should be treated with caution.
2. The impact of the intervention did not appear to be affected by pupils' eligibility for the Early Years Pupil Premium.
3. The trial found suggestive but inconclusive evidence that Maths Champions improved the quality of maths provision in participating nurseries.
4. A third of nurseries taking part in Maths Champions were very or partially engaged with all core features of the intervention.
5. Most nurseries were positive about Maths Champions and its impact on settings and children. However, some also raised the burden on nursery staff time as a potential issue.

EEF security rating

The findings from this study have low security. This was a well-designed randomised controlled trial. However, the security of the trial was compromised by the fact that 36% of the pupils who started the trial were not included in the final analysis. This was largely caused by children leaving participating nurseries or not attending nursery on the day of the post-test. This was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools.

Additional findings

Children who participated in Maths Champions made the equivalent of two additional months' progress in their language development, on average, compared to children in the control group. This result also has low security and should be treated with caution. Neither the impact on maths nor the impact on language were statistically significant. This means that, in this trial, even if the intervention had not had an impact, the probability that just by chance we would have observed an effect size as large as the one found is greater than 5%. The effect of the intervention was not substantially different between girls and boys, nor was it changed by the number of hours a child attends nursery.

82% of nurseries in the intervention group were at least minimally engaged with the intervention. A third of nurseries were very or partially engaged in all of the core aspects of the intervention. If staff were not engaged in the project, it was often due to apprehension about workloads and assessments, low levels of maths confidence, or lack of time to focus on the programme. Nursery staff reported that professional development was an important motivation for starting the project and most staff reported that the training was useful. Staff also reported improvements in their confidence in teaching maths. Participants in the project were particularly positive about the resources provided by the programme. 71% of practitioners and 89% of Maths Champions viewed the resource bank as 'quite' or 'very' useful.

Cost

Maths Champions cost each nursery around £216 per year, or £9 per pupil per year when averaged over three years. It costs around £478.60 for the first year and £85 for each subsequent year. The main financial costs were those associated with the training and on-going support. On average, Maths Champions spent over three hours per week on the programme, with half of them doing this work in their own time. The workload was lower for the other early years practitioners involved, and 22% of them reported that they completed their professional development work outside of their work hours.

Table 1: Summary of impact on primary outcome

Primary outcome	Effect size (95% confidence Interval)	Estimated months' progress	EEF security rating	No. of pupils	P value	EEF cost rating
Mathematics (ASPECTS)	0.10 (-0.13, 0.33)	2	🔒🔒🔓🔓🔓	628	0.41	**£** £ £ £ £

Introduction

Intervention

Intervention Description

Maths Champions is a one-year programme developed by the National Day Nurseries Association (NDNA) with the aim of improving the knowledge, skills and confidence of nursery practitioners in order to improve the quality of maths provision within their setting, and ultimately improve early mathematics achievement. The Maths Champions programme, itself, and most associated resources and training, are based online. Within each nursery, the Maths Champions programme is delivered by a dedicated "Maths Champion" who is a graduate practitioner (being a university level graduate is a prerequisite to being a Maths Champion). Using the programme resources and training available online, the role of the Maths Champion is to support non-graduate practitioners within the setting to evaluate their current practice, assess their current level of mathematical knowledge and skills, develop knowledge and understanding, and build confidence in teaching maths.

The programme starts with two online courses to support the setting's graduate in becoming a Maths Champion. These are intended to help the Maths Champion evaluate and measure the impact of their practice, and give them the skills to mentor and lead the team in the programme. Each course takes around two hours. Once completed, the Maths Champion then works with a team of the other non-graduate (in most cases) practitioners within their settings; whilst not everyone in the setting needs to participate, it is essential for this trial that those practitioners working with children aged three or above do, as this is the target group of the evaluation.

The Maths Champion then evaluates the nursery's current practices using audit tools provided online. This includes tools to gauge:

- Staff confidence: self-reported confidence levels in teaching Maths using a questionnaire.
- Current staff mathematical ability: Through practitioner self-assessment, using the Basic and Key Skill Builder (BKSB) diagnostic tool (or equivalent) to identify the areas of strength and difficulties within each practitioner's mathematical knowledge and understanding. Other practitioners are to be supported by the Maths Champion throughout the course of the intervention with the development of a practitioner action plan, to develop their mathematical skills, and complete sections of the BKSB course online throughout the year.
- Current child ability: using a tracker for staff to record how many children in their groups are at the emerging, expected or exceeding stage of maths in relation to the Early Years Foundation Stage (EYFS) Development Matters statements.

The purpose of the audit is to identify to the Maths Champion the areas of the maths learning environment that need improvement from which they can develop their setting's 'action plan'. The audit identified areas for improvement for the Maths Champion to focus on. For example, improvements could be either practice related or environment related. Maths Champions focused on these improvements during the project. The NDNA project team provided support for these improvements via one to one monthly calls reviewing progress to planned actions. Throughout the course of the year, the Maths Champion and the other practitioners will be encouraged to utilise the online resources, which are focused around activities to incorporate numeracy concepts within play and within interactions. The Maths Champion has a log-in to the Maths Champion's website that allows them to access all parts of the programme including:

- An online resource bank providing math ideas to build into the nursery's daily practice. These resources are reinforced by short monthly webinars with focused themes, to help the Maths Champion and other practitioners implement them into their everyday practice. Live webinars are run each month, but are recorded so can also be watched at a later date. Monthly webinars touched on some of the resources, for example the core activities and the resources that had to be completed during the programme.

- Three further online courses throughout the year for the Maths Champion to complete, which they are expected to disseminate to their team. These focus on:

  (1) Understanding how to observe, assess and plan maths sessions and connect it to other aspects of learning.

  (2) Teaching numbers to different ages and stages of progression.

  (3) Teaching shape, space and measure to different ages and stages of progression.

- Access to the Basic Key Skills Builder (BKSB) maths skills assessment tool which aims to help the team in each setting improve their own personal maths skills with 1-1 support from the setting's Champion (taking around 10-15 minutes per month).

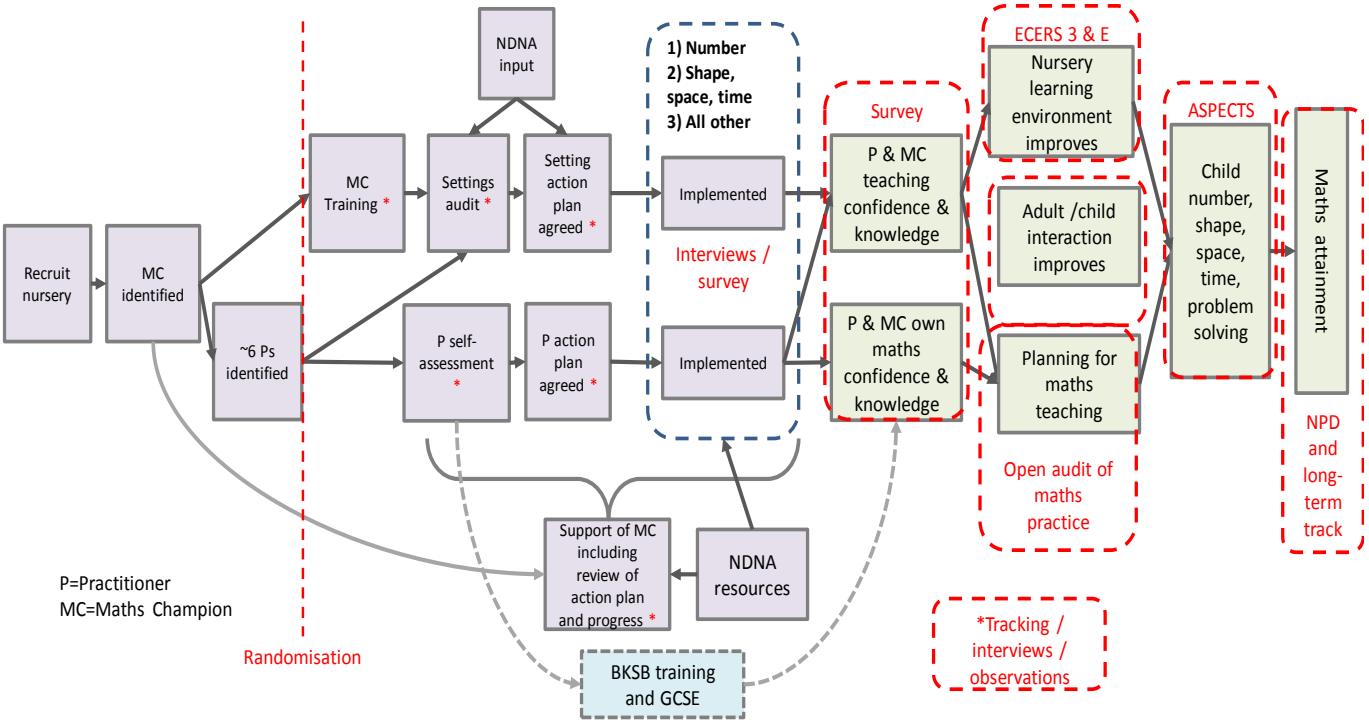
Throughout the programme the Maths Champions also receive tailored 1-1 support from NDNA.

Towards the end of the programme, an outcomes audit is conducted by the Maths Champion. This includes the reassessment of staff confidence levels, staff mathematical skills (through the BKSB), settings and resources audit, and child ability.

A TIDieR (Template for Intervention Description and Replication) checklist is provided in Appendix C.

The Maths Champions Intervention: Theory of change is provided in Appendix D

Figure 1: Maths Champions Logic Model



Audit of Intervention

The independent evaluation of Maths Champions included a pilot phase and effectiveness trial. Further details of the methods of the evaluation are provided below. In parallel with the pilot study, the University of Oxford conducted an audit of the Maths Champions intervention (between October 2015 and June 2016), resulting in the creation of a Theory of Change model (Appendix D). The University of Oxford suggested changes to the intervention based on the findings of the audit, which included reorganising the resources into useful and meaningful categories, clarifying the elements of the programme which were mandatory so that there would be consistency across nurseries with delivery and enhancing the materials so that they better focused on understanding of maths through play. As a result of this feedback, NDNA reorganised and created new resources for the resource bank, particularly focused around the order in which children learn, and number sense (the order in which children learn to identify separate items then start to be able to count them using one to one correspondence). A greater focus on this was introduced to the webinars. Additionally changes were made to the main webpage accessed by Maths Champions so that some elements were mandatory to complete before getting access to the next step or training.

A further change to the intervention based on the audit was the introduction of an Action Research element. This requires the Maths Champion to identify one area in their maths provision to focus further on, research approaches for development (and reasons for doing this), plan, execute and evaluate the impact. In the pilot study, Action Research was part of the resource bank and not highlighted as a compulsory component of the intervention. However, based on the pilot phase NDNA felt this was an important part of the intervention and had a more significant impact than had originally been expected; therefore, for the main trial, it was proposed that it should be considered as a core element with NDNA providing guidance and support to Maths Champions on this element as part of their one to one support around the action plan.

In addition, the University of Oxford team conducted some phone interviews with previous Maths Champions and provided feedback to NDNA based on these. Ten interviews with managers of early years settings in the private and voluntary sector were carried out during February and March 2016. The sample was selected by the NDNA. The report produced by the University of Oxford team did flag up some other areas for improvement which were implemented as a result.

Process Evaluation/Assessment of Implementation Fidelity

Although the intervention is designed to be a one year intervention, during this evaluation most nurseries allocated to the intervention group delivered Maths Champions between September/October 2016 and June/July 2017. The process evaluation of Maths Champions identified that during the evaluation period some face-to-face support was provided by NDNA to nurseries implementing Maths Champions (this has not been standard practice when the programme has been delivered outside of this trial). Five regional induction sessions were held, which the Maths Champion at participating nurseries was invited to attend. NDNA staff members also visited 24 settings (of the 54 allocated to the intervention group) to help support the implementation of Maths Champions. NDNA chose to visit nurseries they thought may be at risk of dropping out of the intervention, some nurseries they approached refused visits. They also visited nurseries where they had identified good practice (in order to get information for internal case studies). Fidelity of implementation is discussed in further detail below; 18 nurseries (of the 54 allocated to the intervention group, 33%) were considered to have implemented Maths Champions as intended.

Background evidence

A large proportion of children in the UK (e.g. 13% in England) do not meet the expected levels in mathematics by the end of primary school (Department for Education, 2015). Research suggests that early mathematics achievement is correlated with both achievement in mathematics and general educational attainment in later life (e.g. Duncan et al. 2007; Jordan et al, 2009). It has been suggested that early interventions are the most cost effective and efficient approach to improving children's outcomes (across a variety of health and education outcomes) (Easton & Gee, 2012), emphasising the importance of early maths intervention. Attendance at pre-school and quality of pre-school provision have been shown to predict children's scores on maths and reading assessments at Key Stage 1 (Sylva et al. 2004), maths and science achievement at Key Stage 2 and 3 (Sammons et al. 2011) and quality of GCSE results (Sylva et al. 2014). Sylva et al. (2014) also found that quality of pre-school had a stronger influence on GCSE Maths and English scores in children whose parents had lower qualification levels, than those whose parents had higher qualification levels. This indicates that high quality early numeracy education at pre-school can have long lasting effects which may help to narrow the gap in achievement throughout life.

Working within this context, the Maths Champions one-year programme was developed by NDNA with the aim of improving the knowledge, skills and confidence of nursery practitioners in order to improve the quality of maths provision within their setting, and ultimately improve early mathematics achievement. A two-year pilot of the programme funded by the Department for Education (DfE) indicated that the programme increased practitioner confidence and skills in delivering numeracy in nurseries (NDNA 2014); however, as this research did not have a comparison group, further robust evidence was needed to strengthen the evidence base. Consequently, the Education Endowment Foundation (EEF) commissioned the University of York and Durham University to conduct an independent evaluation of the Maths Champions intervention.

As part of this independent evaluation of Maths Champions a pilot study was designed and conducted from January 2016 until April 2016 to investigate the feasibility of an effectiveness trial and to pilot possible trial processes. During this phase it was also planned that the Maths Champions programme would be audited in line with current evidence-based guidance on teaching early years' maths (e.g.

Frye et al., 2013; Williams, 2008). This would lead to further development of the intervention and a theory of change model, allowing the programme to be specified more clearly in readiness for a definitive effectiveness trial.

The quality of nursery provision and levels of staff qualifications within Private Voluntary and Independent (IPV) settings have been found to be lower than within integrated centres and schools (Melhulish & Gardiner, 2017); with many nursery practitioners in PVI settings not having higher than level 2 qualifications (APPG Maths & Numeracy, 2014). The PVI sector were also found to be less well equipped to provide equal quality education to disadvantaged and advantaged children (Mathers & Smees, 2014). With this in mind, the effectiveness trial was designed to recruit PVI nurseries only, with the rationale that the Maths Champions programme might support and enhance the knowledge of staff in these settings to be able to deliver higher quality numeracy education.

Evaluation objectives

Pilot Study

The primary aim of the pilot study was to consider the feasibility of conducting an effectiveness trial through piloting aspects of the research methodology, with particular regards to the eligibility, consent and follow-up of children within PVI nursery settings.

Specific objectives were to:

(1) Estimate how many children within each setting fulfil the eligibility criteria in order to provide data for a sample size calculation/recalculation.

(2) Develop a three-tier parental/carer consent process and to estimate the proportion of parents who are willing to provide opt-in consent for:

(i) Their child to participate in the baseline and outcome assessments;

(ii) The evaluation team to contact them by telephone or email to collect school destination data;

(iii) The linking of their child's data to the NPD.

(3) Pilot training for nursery staff in the collection of child attainment data using the proposed primary outcome for the effectiveness trial (ASPECTS; Assessment Profile on Entry for Children and Toddlers).

(4) Pursue a National Pupil Database (NPD) request using data available (child name, date of birth, school destination, home postcode) to estimate the proportion of children for whom long-term educational outcomes could be collected.

(5) Trial the intervention feedback survey to inform the process evaluation

In parallel to the work of the evaluation team, Sandra Mathers and Maria Evangelou at Oxford University were to audit Maths Champions in line with the current evidence-based guidance of teaching early years' maths (e.g. Frye et al., 2013; Williams, 2008), and support further development of the intervention and a theory of change model (Appendix D).

Effectiveness Trial

The research questions of the effectiveness trial impact evaluation were:

- What is the impact of the Maths Champions intervention on the mathematical development and skills of children aged three and four years? [Primary outcome]
- How effective is the Maths Champions intervention at improving nursery practitioners' confidence in supporting children's maths development? [Secondary outcome 1]
- What is the impact of the Maths Champions intervention on the mathematical practice of settings as evaluated using ECERS 3 and ECERS E (Maths)? [Secondary outcome 2]
- What is the impact of the Maths Champions intervention on the language development and skills of children aged three and four years? [Secondary outcome 3]. (NB. The language component of the ASPECTS assessment has been analysed as a single outcome, instead of splitting it into literacy, and phonological awareness outcomes as was originally proposed. This is based on advice from the developers of ASPECTS in order to retain the reliability of the assessment. Secondary outcomes 3 and 4 as listed in the protocol have therefore been merged.)

The process evaluation explored the ways in which the Maths Champions programme was implemented across settings, and the perceived impacts and outcomes of its implementation. The specific research questions were:

- To what extent are the nurseries involved engaging with and delivering the intervention?
- To what extent have nursery practitioners changed their practice from the beginning to the end of the intervention?
- Is fidelity to the intervention being maintained? (NB. The original research question: 'How is the intervention disseminated within the nurseries to other staff?' was amalgamated with 'Is fidelity to the intervention being maintained?' since dissemination to other staff was considered as an element of fidelity as opposed to a discrete focus. Research questions 3 and 5 as originally listed in the protocol have therefore been merged.)
- What are the different stakeholder viewpoints on the intervention?
- What are the key success factors required for the Maths Champions intervention to work well?
- What are the barriers to successful delivery of the intervention?
- What areas of the programme could be further developed following completion of the project?

All research questions were detailed in the 'Independent Evaluation of the Maths Champions Programme: Trial Protocol, Version 3, dated 07/07/2017' (https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Round_8_-_Maths_Champions_AMENDED_2.pdf).

Ethical review

Ethical review was undertaken by Durham School of Education Ethics Committee and was given on 16th March 2016

- Subsequent amendments were submitted and approved as follows:
- Amendment to parent information sheet, approved 23rd June 2016
- Addition of gaining parental consent for researcher to assess children at baseline, approved 20th September 2016
- Submission of Practitioner Interview Information Sheet and Consent Form, approved 13th October 2016
- Submission of Follow-on parent information sheet for ASPECTS post testing, approved 10th February 2017
- Submission of Observational Visit Information Sheet, Maths Champions audit questionnaire and Practitioner confidence survey, approved 30th March 2017

Chairs Action from the University of York Health Sciences Research Governance Committee was also given on 18th April 2016. The committee was notified of relevant subsequent amendments and the approval given by Durham School of Education.

Consent Procedure

Agreement to Participate was sought from the Nursery Setting Manager for nurseries to take part in the evaluation. Parental opt-in consent was sought for: conducting baseline and outcome assessments on children, contacting parents to collect school destination data, and linking of child's data with the NPD. Opt-in consent was sought from nursery practitioners for observations, interviews and surveys.

Project team

The Maths Champions intervention was developed and delivered by NDNA:

Stella Ziolkowski, Director of Quality and Workforce Development – Overarching project lead, NDNA

Madeleine Robinson, National Training Manager – Project manager, NDNA

Jo Baranek, Lead Early Years Advisor – Maths Champions Lead, NDNA

Fiona Bland, Early Years Advisor – Maths Champions Lead, NDNA

Kathryn Moses – Project Co-ordinator, NDNA

Sandra Mathers and Maria Evangelou from the University of Oxford acted as advisors to NDNA providing feedback on the intervention at the pilot stage.

The Evaluation team included members of the York Trials Unit, University of York, panel of evaluators, and members of the Durham University panel of evaluators. The York Trials Unit was the contractual lead on the project.

ASPECTS assessment was conducted by evaluation team members at Durham University.

ECERS assessment was conducted by A+ Education Ltd – Clare Williams, Director (during the ECERS assessment, Sandra Mathers, University of Oxford, was director of A+ Education Ltd).

Evaluation Team

Hannah Ainsworth (York Trials Unit, University of York)

Joint principal investigator of this evaluation. Her role included overall oversight of the impact, process and cost evaluation elements, contributing expertise to the design and conduct of the evaluation as well as to the writing of the final report. Hannah was on maternity leave from November 2015 – November 2016.

Vic Menzies (School of Education, Durham University)

Joint principal investigator of this evaluation. Her role included oversight of the evaluation work done at CEM, contributing expertise to the design and conduct of the evaluation and leading on the process evaluation. Vic was on maternity leave from October 2016 until November 2017.

Dr Lyn Robinson-Smith (School of Education, Durham University)

Co-investigator and acting joint principle investigator during Vic's maternity leave as well as some months before. Her role included contributing to the design and conduct of the evaluation, as well as leading the training and coordination of staff to conduct the baseline and outcome assessments. Lyn undertook all trial management responsibilities and oversaw the process evaluation. Lyn was on maternity leave from September 2017 until completion of the evaluation.

Professor David Torgerson (York Trials Unit, University of York)

Co-investigator and acting joint principle investigator during Hannah's maternity leave. His role included supporting the design and conduct of the evaluation including the economic evaluation and contributing to the write up of the final report.

Gemma Stone (School of Education, Durham University)

Covered joint principle investigator role during Vic and Lyn's maternity leave. Gemma contributed to the process evaluation and writing of the final report.

Professor Catherine Hewitt (York Trials Unit, University of York)

Co-investigator of this evaluation, providing input and support into the trial statistical analysis.

Caroline Fairhurst (York Trials Unit, University of York)

Co-investigator of this evaluation undertaking the randomisation and statistical analysis, and contributing to the write up of the final report.

Professor Carole Torgerson (School of Education, Durham University)

Co-investigator of this evaluation, contributing expertise to the design and conduct, as well as to the writing of the final report.

Louise Elliott (York Trials Unit, University of York)

Louise contributed to the conduct of the trial and to the final report.

Louise Gascoine (York Trials Unit, University of York)

Louise was involved in the pilot study and conducted trial coordination responsibilities, including trial registration and development of the practitioner survey and contributed to the writing of the final report.

Dr Kerry Bell (York Trials Unit, University of York)

Kerry conducted trial coordination responsibilities. Kerry was on maternity leave from September 2017 until completion of the evaluation.

Sarah Hallett (School of Education, Durham University)

Sarah was the evaluation administrator facilitating the administration of ASPECTS and liaising with settings for visits for data collection purposes. She will also assist with the collection of school destination data from parents.

Jess Hugill (School of Education, Durham University)

Jess contributed to the process evaluation analysis and report writing.

We would also like to thank all the independent assessors who completed ASPECTS assessment and ECERS assessment.

Trial registration

Upon agreement of the trial protocol the trial was registered with ISRCTN in May 2016 and given the following number ISRCTN13051035

The trial was listed as retrospectively registered as informal nursery recruitment for the effectiveness trial had already begun; however, no children had been recruited for the effectiveness study at the point of trial registration and randomisation had not been conducted.

Pilot Study Methods

Pilot Study design

The pilot study involved piloting trial processes and procedures relating to recruitment, consent, outcome testing, and follow up in six PVI nurseries. The pilot study took place between February and April 2016

Following PVI nursery recruitment and gaining parental consent for children's participation, each participating nursery was requested to select one member of staff (the 'Maths Champion') to assess all children using ASPECTS and to participate in necessary training administered by the evaluation team via webinars. Within each nursery, the ASPECTS assessment was conducted on a random sample of ten children who met the eligibility criteria and whose parents had provided consent for testing, in order to keep the testing burden on nurseries to a minimum.

The evaluation team (at Durham) developed and held training webinars for nursery staff to learn about the ASPECTS assessment. Three webinars were run on the 27th May, the 7th June and the 10th June 2016. These were held at 1pm and lasted for between 40 minutes and 1 hour depending on the number of questions from participants during the session. Training included:

- What the ASPECTS assessment covered
- How to install the ASPECTS programme on a computer/laptop
- How to use the programme and conduct the assessment
- A demonstration of running the ASPECTS assessment
- How to return the ASPECTS data to CEM for processing
- How to interpret the data

To gauge the quality of the training webinar, a member of the research team visited each setting to observe the assessment being delivered. During the visit to the nursery, the researcher also requested the Maths Champion and one other practitioner within the setting complete a survey regarding their experiences of the Maths Champions programme (Appendices E and F). The survey was designed to take no longer than 15 minutes to complete and would help inform the development of the survey to evaluate Maths Champions in the main effectiveness trial.

Nurseries were contacted via telephone from September 2016 to collect necessary pupil data (school destination) of children whose parents provided consent for long-term follow up. In the event that these data were unavailable at the nursery level, the evaluation team contacted parents/carer via email or telephone, where consent for such contact had been given.

Participant selection

Nurseries

The pilot phase included PVI nurseries in which the NDNA's Maths Champion programme was already established during the 2015/2016 academic year. It was estimated that there were approximately 60 PVI nurseries that were currently implementing Maths Champions in the UK. The pilot study required the participation of six nurseries in total. NDNA made initial contact with the nurseries inviting a sample of them to participate in the pilot study using an information sheet provided by the evaluation team (Appendix G). Nurseries interested in participating in the pilot study were requested to contact the

evaluation team for further information or to complete, sign and return the Memorandum of Understanding (MoU) for the pilot trial (Appendix H) to the evaluation team. Participation in the pilot study was on a first come first served basis on the return of a signed MoU. For participating in the pilot study, nurseries received a £250 Amazon voucher.

Participants

Each participating nursery was requested to provide the evaluation team with the number of children within their setting who fulfilled the pilot study's eligibility criteria: being three years old at the start of the intervention; due to attend school in September 2016; and attending nursery for a minimum of 15 hours per week. Following this, the evaluation team sent each nursery 'parent information letters' (Appendix I) and 'consent forms' (Appendix J) to distribute to parents of children who met the eligibility criteria. Parents were requested to return completed consent forms to nursery staff who were advised to post the consent forms to the evaluation team.

Outcome Measures

1. Number of PVI nurseries recruited
2. Number of eligible children identified
3. Number of children with parental consent at each of the three levels
4. Feedback from observations of nursery practitioners conducting ASPECTS assessment
5. Feasibility and appropriateness of the practitioner survey

Sample Size

The aim was to recruit 6 PVI nurseries to take part in the pilot study to provide information to inform the trial design and implementation of the effectiveness trial.

Results

Nursery recruitment

In total, the evaluation team received eight Memorandum of Understandings (MOUs) via NDNA from nurseries willing to participate in the pilot study. One nursery was deemed ineligible as it had been running Literacy Champions (for more information see: https://www.ndna.org.uk/childcare-training-literacy-champions-programme) within their setting, and another was kept as reserve in case of drop-out; however, their assistance was not required. So six nurseries participated in the pilot.

Research question 1: *Gauge how many children within each setting fulfil the eligibility criteria, that is, being three years old at the start of the intervention, due to attend school in September 2016 and attend nursery for a minimum of 15 hours per week.*

All six nurseries provided information regarding the number of children within their setting who met the relevant eligibility criteria. The number of children in each setting who met eligibility criteria ranged from 10 to 71, with an average of 27 per nursery.

Research question 2: *Develop and gauge response to a three-tier parental/carer consent process to estimate the proportion of parents who are willing to provide opt-in consent for:*

> *(i) Their child to participate in the baseline and outcome assessments (ASPECTS, described below)*
>
> *(ii) The evaluation team to contact them by telephone or email to collect school destination data*
>
> *(iii) The long-term tracking of their child's education data post-nursery using the NPD.*

In total, 4/6 nurseries participating in the pilot study returned the parent consent forms as requested by the evaluation team. Of the 2/6 nurseries who the evaluation team did not receive consent forms from, both stated they posted them to the evaluation team and that they must have got lost in the post. The number of parent consent forms received ranged from five to 13, with an average of eight per setting. The total number of consent forms received by the evaluation team was 32 and all parents indicated they were willing to opt-in to all the three elements.

Research question 3: *Pilot training for nursery staff in the collection of child attainment data using ASPECTS.*

The training session ran for approximately 30 minutes and was attended by five staff from 5/6 of the nurseries. One staff member was unable to attend the live webinar; however, they were sent the webinar recording to view at a more convenient time. To gauge the quality of the training webinar, a member of the research team visited each setting to observe the assessment being delivered. These observations indicated that the training was well received and at the time of the visit to each participating nursery, all practitioners had installed the software onto their system ready to use. Overall, observations indicated that the training ensured replicability of data collection across the settings, with one exception relating to asking children to write their name. Name writing is the first activity ASPECTS requests children to complete. Children are provided with a pencil and a piece of paper and asked to write their own name which the assessor then scores against five criteria. On observation, one setting provided children with name cards for them to copy. It was recommended that clarification on this should be added to the training of practitioners/external testers within the main trial to ensure consistent data collection throughout settings. In total, 30 ASPECTS scores were gathered across the four participating, consented nurseries (ranging from five to 13 per nursery).

Research question 4: *Pursue an NPD request using data available (child name, date of birth, school destination, home postcode) to gather and track long-term educational outcomes.*

The pilot study did not require the long-term tracking of participating pupils' educational outcomes. Rather, the feasibility of long-term tracking and the matching process (gathering relevant pupil level data to enable linkage to their unique pupil number (UPN)) to the NPD was explored. This was necessary as children in PVI nurseries are not assigned a UPN until they reach primary school; however, the child's name, date of birth and school destination should be sufficient for linkage to the NPD. Exploring how many children could be matched to the NPD in the pilot study aimed to provide an estimate of the proportion of children that could be matched in the main trial in order to collect long-term educational outcomes in the future.

Nurseries of the 32 children whose parents provided necessary consent were contacted to obtain school destination data (child's forename and surname, date of birth and home postcode were collected on recruitment). School destination was gathered directly from the nursery in 28/32 cases with 3/32 being unavailable (in 1/32 the child had not yet moved to school and remained in the nursery). Of the 3/32 cases where nurseries were unable to provide school destination, the evaluation team collected this information from parents/guardians successfully in one case; however, the remaining two parents/guardians were unresponsive to contact.

The NPD were able to provide the evaluation team with the number of UPNs which could be matched with UPNs found within the NPD. If a UPN could not be matched it was because either (1) the child(ren) could not be found in the NPD based on the information provided (no match), or (2) the child is located within the NPD (match) but a UPN has not yet been recorded for the child by the school in their School Census returns. The NPD was successfully able to match 30 children (94%) of the 32 provided by the evaluation team.

Research Question 5: *Trial the intervention feedback survey to inform the process evaluation.*

The intervention feedback survey was completed by four from the six participating nurseries (correlating with those that had collected consent for children, as described above). Within each of these four nurseries, all nominated Maths Champions completed surveys, and between one and three other practitioners.

In the pilot, a single survey was piloted that covered both the nominated Maths Champion and other practitioners. Feedback and review of responses highlighted that several questions in the survey were only pertinent to the Maths Champion themselves (questions focusing on elements of the programme that only the Maths Champion had access to and was required to complete). As such, two separate surveys were produced for the main trial, to ensure that only relevant questions were asked to each group.

Conclusions

The pilot study indicated that nurseries would be willing to implement Maths Champions and that the input required of nurseries taking part in a randomised evaluation (identification of eligible children, informing parents/carers about the evaluation, collecting completed consent forms, and supporting pre and post-test data collection) would be acceptable and possible for nurseries to undertake. Since the pilot study was not a randomised evaluation, the willingness of nurseries to be randomised and to adhere to randomisation procedures could not be inferred.

Nurseries participating in the pilot study implemented Maths Champions as intended and no issues were raised regarding implementation. Informal feedback and information collected through surveys about the intervention was positive. The audit conducted by the University of Oxford during the pilot study resulted in some recommendation for changes to the Maths Champions intervention in the main trial. These changes have been detailed below and in the 'Intervention Description' section of this report.

On average, nurseries identified 27 eligible children per nursery, and consent was gained for, on average, eight children per nursery. All consenting parents/carers provided consent for all three elements of the evaluation.

Attendance at live or recorded webinars was sufficient to support nursery practitioners in collecting ASPECTS assessment. It was identified that clarification on how to get children to write their name was required.

Pursuing a NPD request following the main trial will be worthwhile since an estimated 94% of children may be able to be matched to enable long term follow up, based on the percentage matched in the pilot study.

Recommendations for the main effectiveness trial

- In relation to pupil recruitment, increased communication between evaluation team and nurseries and between nurseries and parents/carers, will be required in order to meet the necessary sample size within each nursery, since the pilot study demonstrated an average recruitment of 8 pupils per nursery rather than the desired 10 pupils.
- ASPECTS training webinars should be recorded. ASPECTS assessors should attend live training sessions where possible, and view recorded training sessions at a later, more convenient time if not.
- Clarification on instructions for getting children to write their name (which forms part of the language assessment of the ASPECTS assessment) should be added to the training of

practitioners/external testers within the main trial to ensure consistent data collection throughout settings.

- Intervention feedback surveys should be separated into two separate surveys to capture responses from the nominated Maths Champion and other practitioners separately to reduce burden on respondents.

Effectiveness Trial Methods

Trial Design

A pragmatic, two-armed, cluster randomised controlled trial was conducted. Nurseries were randomly allocated 1:1 to one of two groups:

- Intervention – nurseries allocated to receive the one-year Maths Champions intervention; or
- Control – nurseries allocated to continue with usual nursery provision.

Changes from the Protocol

An amendment to the Protocol was made on 17 July 2017 to clarify that post-testing would be conducted by the evaluation team (rather than sub-contracted).

No other design changes occurred from the original agreed protocol.

Participant selection

Nursery Recruitment and Eligibility

Nurseries were recruited by NDNA, supported by the evaluation team at Durham and York. NDNA ceased general registration for the Maths Champions intervention in December 2015 to ensure capacity would be sufficient to support the project. NDNA began recruitment for the trial in February 2016 with regional events. A targeted email was sent to approximately 1000 settings in local authority areas with high levels of deprivation. The evaluation was also mentioned in a newsletter to NDNA members, reaching approximately 5000 settings. Other avenues for publishing the evaluation included; a press release in Nursery News (NDNA's membership magazine); use of social media; information added to NDNA webpage; and flyers produced for use at events. Nurseries interested in taking part were asked to complete an online expression of interest form. At this point nurseries were screened against the eligibility criteria (provided below) and those meeting the eligibility criteria were provided with further detailed information and asked to sign the Memorandum of Understanding (Appendix K) if they wanted to take part. All nurseries (both intervention and control) were offered £500 for taking part in the research; nurseries were able to invoice NDNA directly at the end of the study on completion of the post-tests.

Nursery Eligibility Criteria:

- PVI nurseries whose child population included children who are three years old.
- Nurseries located in areas of high deprivation (initially nurseries located in the 40 most deprived local authorities [as identified using the IDACI (Income Deprivation Affecting Children Index) September 2015], although this was subsequently extended to other local authorities in order that the sample size could be achieved. As far as possible the evaluation aimed to include nurseries who were providing for disadvantaged children.)
- Nurseries not previously involved with the Maths Champions intervention.
- Nurseries who agreed to all study requirements outlined in the Memorandum of Understanding (Appendix K).

Child Recruitment and Eligibility

Participating nurseries were asked to identify all children in their nursery who met the following eligibility criteria:

Child Eligibility Criteria

- Children who were three years old at the start of the intervention.
- Children due to start school in September 2017.
- Children attending nursery for a minimum of 15 hours per week.

The inclusion criteria relating to age and minimum number of hours were set to ensure that children in the evaluation were in the nursery long enough to receive three academic terms of the intervention.

The evaluation team provided each nursery with hardcopies of information sheets for parents (Appendix L), a document with frequently asked questions (Appendix M), and opt-in consent forms (Appendix N), and requested that they distribute them to the parents/guardians of all eligible children. Whist this was the protocol, there was anecdotal evidence to suggest some nurseries self-selected the parents/guardians in the setting who would be most responsive to their child's participation in the trial, and approached children who were close to meeting the inclusion criteria, e.g. children who were actually two at the start of the intervention but about to turn three. Nurseries were asked to compile any consent forms returned by parents/carers and post them to the evaluation team. Parents/carers were required to give opt-in consent for their child to participate in the baseline and post-intervention outcome testing. Parents/carers were also requested to provide opt-in consent to be contacted should school destination data for their child not be available from the nursery, and also for long term tracking of their child's educational attainment via the NPD.

Nurseries were requested to return all completed parent/guardian opt-in consent forms to the research team via secure post or email. During the early stages of pupil recruitment it became evident that in some settings fewer than 10 opt-in consents were being received (sample size calculations, provided in detail below, assumed 10 pupils per setting). To offset this, the evaluation team increased the maximum number of children to be assessed in nurseries who were yet to return consent forms from 10 to 14.

Following this process, some nurseries were unable to conduct the baseline ASPECTS testing themselves (a variety of reasons were given by nurseries but mostly they related to their own time constraints, capacity, and the additional burden on them) and so additional consent was sought from parents for children to be assessed by researchers from the evaluation team (Appendix O).

Before post-testing a further letter was sent to parents from the evaluation team, via nurseries, reminding parents that they had previously given consent for testing and providing some updated information on data protection (Appendix P).

Outcome measures

Primary Outcome

Maths attainment at the end of nursery measured using ASPECTS, which was produced by the Centre for Evaluation and Monitoring (CEM) at Durham University. Participating children were assessed using ASPECTS at baseline and were followed-up a year later. ASPECTS has been specifically designed for children aged between three and five years old (36 to 60 months), and is aligned with crucial elements of the Early Years Foundation Stage (EYFS) Prime and Specific areas of Learning and Development. The range of possible scores for the mathematics score of the ASPECTS is 0 to 29 with a higher score indicating greater attainment.

Within each nursery, testing was conducted on approximately ten children who met the eligibility criteria and whose parents provided the relevant consent. If there were fewer than 14 eligible, consenting children per setting, all were tested where possible. If there were 14 or more eligible, consenting children then up to 14 were randomly selected for testing. Nurseries had the option to additionally assess

children who were not randomly selected if consent was in place and they had capacity to do so; these data are included in the trial analysis.

At baseline, in most cases, the children were tested by their nursery practitioner who received necessary training administered by the evaluation team via webinar. Where nurseries were unable to conduct baseline testing a research assistant trained by the evaluation team conducted baseline testing.

At follow-up, children were assessed by an independent, blinded assessor trained by the evaluation team.

Limitations of testing in the early years are the young age and temperament of children who may or may not be responsive at the time of testing, potentially leading to missing data. To minimise this issue, we tried to ensure that the child was familiar with the adult who administered the baseline assessment when they are very young (three years old) and at outcome assessment the child could be accompanied by a familiar adult if necessary. It is worth noting however that while an adult who already knows the child may help the child feel more comfortable they may also influence how the assessment is conducted and subconsciously or consciously adjust their scoring of the assessment based on what they know the child can do in other circumstances rather than just at the moment of assessment. For this reason, the outcome assessment was conducted by an independent researcher to reduce potential bias.

Secondary Outcomes

Language (reading and phonological awareness) score from the ASPECTS (child-level) is a secondary outcome. This is scored from 0 to 53, where a higher score indicates greater attainment.

Practitioner confidence and beliefs were gathered via a short survey, adapted from the version used by Chen et al (2014) (Appendix Q). We requested this should be completed by the Maths Champion and all other practitioners who work with children aged 3 and above. The survey was completed on paper at post-intervention only and consists of 28 items in total over three subscales described below in table 2. Each item was scored on a Likert scale from strongly agree (1 point) to strongly disagree (5 points). Some items are reverse scored. Scores for items in the subscales were summed to produce summary scores for each subscale. The three subscales represent three different constructs and so it is not possible to combine them into a total score.

Table 2. Practitioner confidence and beliefs measure subscales

Subscales	Number of items in scale	Range of possible scores
Beliefs About Nursery Aged Children and Maths	8	8-40
Confidence in Helping Nursery Aged Children Learn Maths	11	11-55
Confidence in Own Maths Abilities	9	9-45

Nursery environment/provision measured using the Early Childhood Environmental Rating Scales 3 (ECERS-3) and the Early Childhood Environmental Rating scale extension (ECERS-E) was collected at the nursery level at the end of the trial in both intervention and control nurseries. A full day of observation (6 hours) was completed in each setting by a subcontracted organisation (A+ Education Ltd) who subsequently provided the data to the evaluation team for analysis. Assessors were blinded to group allocation.

The ECERS-3 and E are designed to give a snapshot of provision within a setting on a particular day. ECERS-3 consists of 35 items organised into six subscales: (1) space and furnishings, (2) personal

care routines, (3) language and literacy, (4) learning activities, (5) interaction, and (6) programme structure. Each item is scored from 1=inadequate to 7=excellent. Four items of the learning activities subscale focus specifically on maths provision: (23) maths materials and activities, (24) maths in daily events and (25) understanding written numbers. The whole ECERS-3 was completed for each nursery during a 3-hour morning observation, enabling an overall mean score (1-7) to be generated (the sum of all items divided by number of items scored). The three maths items were also completed a second time during the afternoon, enabling creation of a maths sub-score based on practice across the whole day (the ECERS-3 maths item sub-score). The aim was to ensure that no information on maths practice was lost.

The ECERS-E comprises four subscales (literacy, maths, science/environment and diversity), of which only the maths subscale was completed. The maths subscale comprises four items: (1) counting and application of counting, (2) reading and representing simple numbers, and (3) shape OR (4) sorting, matching and comparing. Each item is scored from 1=inadequate to 7=excellent. A summary maths subscale score is computed by summing the item scores for (1), (2) and the higher of (3) or (4) and dividing by 3 to obtain a total mean score (1-7). The ECERS-E was completed across the full day, as per the author's guidance.

In all, three separate measures were generated by the observations:

1. Overall ECERS-3 mean (general practice during morning sessions)
2. Overall mean of ECERS-E maths subscale (maths practice across the whole day)
3. A maths composite score: the sum of the ECERS-3 maths item sub-score and the ECERS-E maths subscale (maths practice across the whole day)

Other important information

Nurseries were asked to provide data, for each child, on their date of birth, gender, early years pupil premium (EYPP) status, ethnicity and average number of hours the child attends the nursery per week. Some of these descriptive items were collected at baseline, others were collected at the post-testing time point. We acknowledge that the number of hours a child attends nursery could change over the year and be different at post-test than at baseline; however, such changes should be comparable between the intervention and control group.

Compliance and fidelity

Each nursery in the intervention arm was assessed for their implementation fidelity (the extent to which they implemented Maths Champions as intended by NDNA). The NDNA rated each setting on eight core aspects of the intervention including attendance of Maths Champion at compulsory courses, attendance at webinars, use of action plans, etc. on a scale of 2=very engaged ('green'), 1=partially engaged ('amber'), and 0=not engaged ('red') (Table 3). The ratings on these eight components were summed out of a possible 0-16. To investigate the effects of compliance, the nurseries were categorised according to their level of engagement with the intervention in two ways, for the analysis, as follows:

- engaging at least minimally with the intervention (nursery rated by the NDNA as being very or partially engaged in *at least one* of the core aspects of the intervention) vs. no intervention (control nurseries plus all intervention nurseries for whom all core components of the intervention were rated red); and
- good fidelity with the intervention (nursery rated by the NDNA as being very or partially engaged in *all* of the core aspects of the intervention) vs. no or unsatisfactory engagement (control nurseries plus all intervention nurseries for whom at least one core component of the intervention is rated red).

Table 3: Core Components Fidelity Rating

Core Components	Description	RAG rating
Attendance by the Maths Champion at the seven online training courses	All done and completed	Green = 2
	Between 1 and 6 completed	Amber = 1
	None completed	Red = 0
Attendance by the Practitioner at five of these online training courses	All done and completed	Green = 2
	1 or 2 done but needed to be reminded	Amber = 1
	None done	Red = 0
Attendance at live webinar training sessions	3+ webinars	Green = 2
	1-2 webinars	Amber = 1
	0 webinars	Red = 0
Attendance at recorded webinar training sessions	3+ webinars	Green = 2
	1-2 webinars	Amber = 1
	0 webinars	Red = 0
Completion of the 11 core activities	All done and evidence uploaded	Green = 2
	Some done but needed support	Amber = 1
	None done	Red = 0
Completion of BKSB and follow-up activities	BKSB done and activities done on a regular basis	Green = 2
	BKSB done but no activities	Amber = 1
	No BKSB done	Red = 0
Completion and continued use of an action plan	Action plan done and used as working document throughout	Green = 2
	Action plan done, started to be used but then forgotten	Amber = 1
	Action plan not done/not used	Red = 0
Completion of action research throughout the project	Action research started / completed and worked with staff	Green = 2
	Action research started or planned	Amber = 1
	Action research not started or planned	Red = 0

Sample size

From protocol

We make the following assumptions: an intracluster correlation coefficient (ICC) of 0.19[1] and 10 children per nursery with a pre and post-test correlation of 0.70 (CEM, 2001). Based on 120 nurseries (1200 children) we would have 80% power to show an effect size of 0.20 of a standard deviation between the control and intervention groups, allowing for 10% attrition at the child level.

[1] Based on ICC observed in Every Child Counts (ECC) evaluation. Torgerson C.J., Wiggins A., Torgerson D.T., Ainsworth H., Barmby P., Hewitt C., Jones K., Hendry V., Askew M., Bland M., Coe R., Higgins S., Hodgen J., Hulme C. & Tymms P. (2011). The Every Child Counts Independent Evaluation Report. Department for Education.

At randomisation

The final number of nurseries randomised into the trial was 108 (Intervention 54; Control 54); however, 11 dropped out between randomisation and completion of the baseline testing. In total, 845 children were assessed using ASPECTS at baseline across 97 nurseries. With this number, assuming an average of 9 (845/97=8.7, rounded to 9) children per nursery, an ICC of 0.19, a pre-post test correlation of 0.70 and 10% loss to follow-up, we would have 80% power to detect an effect size of 0.22 between the two arms.

Randomisation

In July 2016, the independent trial statistician, who was not involved in the recruitment of nurseries, used a dedicated computer program, MinimPy (Saghaei and Saghaei, 2011), to randomise nurseries 1:1 to the intervention group or to the control group. Minimisation was undertaken to ensure balance between the groups on number of graduate staff and size of the nursery. Some settings were randomised *after* they had committed to the evaluation in principle but *before* baseline testing. This was due to logistical issues; it was originally hoped that baseline testing would be complete by the end of the academic year 2015/2016 and randomisation could take place immediately after. However, in some settings, baseline ASPECTS was only completed at the start of the autumn term 2016. NDNA needed to know which nursery settings were going to be allocated to the intervention group as soon as possible so they could plan training dates and locations with enough time to begin the intervention as soon as possible in the autumn term 2016, and waiting until after ASPECTS was completed would not have left enough time in some settings. Nurseries were **not** informed of their allocation until they had completed the baseline testing.

Nurseries were allocated in three batches (88 in the first batch, 19 in the second, and 1 in the third). Naïve minimisation with base probability 1.0 was conducted, i.e. 1:1 deterministic minimisation. Naïve minimisation was deemed to be sufficient as the allocations were conducted in batches, rather than prospectively, meaning predictability was not a concern and hence a random element was not required (Altman and Bland, 2005). The following minimisation factors were used:

- Number of nursery practitioners in the nursery who were graduates (2 levels; One graduate, More than one graduate)
- Number of children leaving for primary school in 2017 (2 levels; <22, ≥22 [22 was the median from the first batch of 88 nurseries that were randomised])

In the protocol, we proposed that minimisation based on size of nursery, type of nursery (private, voluntary or independent) and whether they had one or more than one graduates would be used to randomly allocate the nurseries to a trial arm. In practice, all recruited nurseries were of the same type (private) and so 'type of nursery' was not required. We defined' size' of the nursery by the proxy measure of number of children leaving for primary school in 2017.

The final number of nurseries randomised into the trial was 108 (Intervention 54; Control 54).

Analysis

The statistical analysis followed the most recent EEF analysis guidance available at the time (**https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/Analysis_for_EEF_evaluations_REVISED_Dec_2015.pdf**, accessed on 27/02/2017). Analysis was conducted in Stata v15 using the principles of intention to treat, where data were available, including all nurseries and children in the groups to which they were randomised irrespective of whether or not they actually received the intervention. A detailed statistical analysis plan can be found here:

https://educationendowmentfoundation.org.uk/public/files/Projects/Round_8_-_Maths_Champions_SAP.pdf.

Statistical significance was assessed using two-sided tests at the 5% level. Estimates of effect with 95% confidence intervals (CIs) and p-values are provided.

The trial was designed and conducted, and has been reported, to CONSORT standards (Schulz et al, 2010; Campbell et al, 2012). A CONSORT diagram has been provided to show the flow of nurseries and children through the trial.

The number of children identified as eligible for the evaluation, the number for whom parental consent was received, and the number actually tested for ASPECTS at baseline and follow up are reported with reasons for non-participation given where available.

The numbers (with reasons and timings) of losses to follow-up (drop-outs and withdrawals) over the course of the trial (pre and post-randomisation) are summarised. The number of nurseries formally withdrawing from the intervention is also given with reasons.

Primary intention-to-treat (ITT) analysis

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, controlling for raw baseline ASPECTS numeracy score, and nursery-level minimisation factors (number of graduate staff and number of children leaving for primary school in September 2017) as fixed effects. The continuous variables that were dichotomised to use as factors in the minimisation were included in their continuous form in the model.

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.

The effect size for the intervention effect was calculated based on the adjusted mean difference between the intervention and control group and the total variance (between plus within nursery variance), obtained from the multilevel model. These two figures are reported. The effect size and 95% CI was calculated using equations (19) and (20) given in Hedges (2007) for cluster randomised designed analysed via multilevel models and allowing for unequal cluster sizes.

The ICC for the primary outcome is presented, and the correlation between the pre- and post-intervention ASPECTS mathematics scores.

Secondary ITT analysis

The primary ITT model was repeated but adjusting only for group allocation and prior attainment (baseline ASPECTS numeracy score) as fixed effects, and nursery as a random effect.

Interim analyses

No interim analyses were undertaken.

Imbalance at baseline for analysed groups

Nursery and pupil characteristics and baseline ASPECTS scores are summarised descriptively by randomised group both as randomised and as analysed in the primary analysis. No formal statistical comparisons were undertaken on baseline data (Senn, 1994), except to report the differences in pre-test scores (maths and language scores from ASPECTS) between the two groups as a Hedges' g effect size (Hedges, 1981). Continuous measures are reported as a mean and standard deviation (SD), while categorical data are reported as a count and percentage.

Missing data

The amount of missing baseline and outcome data is summarised. Where these data are missing, reasons are provided where possible. These include reasons relating to the pupil (e.g. absent on day of testing), the nursery (e.g. nursery withdrew from the trial) or other (e.g. child assessed but data not returned due to technical error). The impact of missing pre- and post-intervention ASPECTS mathematics score data on the primary analysis is assessed using multiple imputation by chained equations to impute missing data using all variables included in the primary analysis model (i.e. group nursery, allocation, number of graduate staff and number of children leaving for primary school in September 2017).

A 'burn–in' of 10 was used (meaning that the first 9 iterations were discarded to allow the iterations to converge to the stationary distribution before the imputation) and 100 imputed datasets were created. The primary analysis was then rerun within the imputed datasets and Rubin's rules used to combine the multiply imputed estimates.

Intervention fidelity and CACE analysis

A Complier Average Causal Effect (CACE) analysis (Dunn, 2005) for the primary analysis (as described in the section: *Primary intention-to-treat (ITT) analysis*) was conducted using an instrumental variable (IV) approach with randomised group as the IV, to account for compliance/engagement of the nurseries with the intervention. The NDNA provided data on whether a Maths Champion was identified at the nursery, whether they were inducted and attending training, and whether they attended the monthly webinars. All this data has been summarised. Formal CACE analysis for the primary outcome at the child-level to investigate the effects of compliance was conducted in two ways, defining compliance of the nurseries as a dichotomous variable as:

- engaging at least minimally with the intervention (defined as the nursery being rated by the NDNA as being very or partially engaged in *at least one* of the core aspects of the intervention – see Compliance and fidelity section, score of at least 1 out of 16), vs no intervention received at all (control nurseries plus all intervention nurseries for whom all core components of the intervention were rated red, score of 0); and
- good fidelity with the intervention (defined as the nursery being rated by the NDNA as being very or partially engaged in *all* of the core aspects of the intervention i.e. rated green or amber (minimum score of 8 and all components scoring at least 1) - see Compliance and fidelity section), vs no or unsatisfactory engagement (control nurseries plus all intervention nurseries for whom at least one core component of the intervention is rated red).

Secondary outcome analyses

ASPECTS language score

The language score from the ASPECTS was analysed in the same way as the primary outcome as described in the section: *Primary intention-to-treat (ITT) analysis*, with the exception that this model was adjusted for baseline ASPECTS language score, and not the mathematics score.

ECERS assessment

Scores for ECERS-3, the maths subscale of the ECERS-E and the ECERS maths composite score were compared between the nurseries in the two groups using linear regression. There was no baseline value for the nurseries to include in the analysis. It was planned that the models would be adjusted for the minimisation factors (number of graduate staff, number of children leaving for primary school in September 2017, in their continuous form) and for the proportion of paid childcare/teaching staff qualified to Level 3 or above and percentage of non-white British children, which have previously been

seen to be predictive of the maths sub-score of ECERS-E (Mathers et al. 2011). However, the proportion of non-white British children per nursery was not comprehensively collected and it was felt that there was too much missing pupil-level ethnicity data collected from children completing the post-test ASPECTS assessment to accurately calculate a reliable estimate of this, so this covariate was omitted from the models.

Sensitivity analyses for ECERS

During the course of the trial, two nurseries (one intervention and one control) amalgamated. The children in these two nurseries have been analysed for the primary outcome according to the allocation of the nursery they were in at baseline, according to the principles of intention to treat. The ECERS is a nursery-level assessment looking at the setting. Since the two nurseries now share one setting, we proposed that one assessment of the environment be made and the resulting score for the ECERS was assigned to both nurseries for the main analysis. In sensitivity analyses, we repeated the ECERS-3 and the ECERS-E analyses but with the two nurseries in question removed to test the robustness of the results to this possible contamination.

Practitioner confidence survey

Responses to items in the practitioner confidence survey are summarised descriptively by trial arm, just for the 'Maths Champion' of the nursery (where this person could be identified), and for all respondents. The three subscale scores for all respondents together were compared between the two trial arms using separate linear regression models, adjusting for the nursery-level minimisation factors (number of graduate staff and number of children leaving for primary school in September 2017, in their continuous form) and highest qualification in mathematics of the respondent as fixed effects, and including nursery as a random effect.

Subgroup analyses

Subgroup analyses looking, separately, at the number of hours a child attends nursery (as a continuous variable), socioeconomic status (Early Years Pupil Premium), and gender have been undertaken for the primary analysis by including interactions with the allocation term.

Implementation and process evaluation

The implementation and process evaluation research questions were developed by the evaluation team in conjunction with the project delivery team. The intention of these questions was to enable the evaluation to track change in relation to the anticipated impact of the Maths Champions programme identified in the Theory of Change (Appendix D). The research questions also relate specifically to Humphrey's identified dimensions within a process evaluation (Humphrey et al, 2016) (Table 4).

Research question 7 of the process evaluation (What areas of the programme could be further developed following completion of the project?) is not directly aligned with an implementation dimension, instead prompting a review of the available evidence to make recommendations for future implementation of the programme.

Table 4: Implementation and Process Evaluation research questions and dimensions

Research Question	Dimensions of implementation
1. To what extent are the nurseries involved engaging with and delivering the intervention?	Fidelity Dosage Reach
2. To what extent have nursery practitioners changed their practice from the beginning to the end of the intervention?	Quality Participant responsiveness
3. Is fidelity to the intervention being maintained?	Fidelity Adaptation
4. What are the different stakeholder viewpoints on the intervention?	Participant responsiveness Programme differentiation
5. What are the key success factors required for the Maths Champions intervention to work well?	Quality Monitoring of control/comparison conditions
6. What are the barriers to successful delivery of the intervention?	Quality Monitoring of control/comparison conditions
7. What areas of the programme could be further developed following completion of the project?	N/A

Three further questions were originally listed in the trial protocol:

- How is the intervention disseminated within the nurseries to other staff?
- How effective and appropriate are the level of support and training?
- What is the acceptability of the intervention and does this differ depending on the experience or qualifications of the staff involved?

However it was decided that the first two research questions above were being implicitly captured by the other seven research questions and as such when results are provided, they are provided against the research questions 1 to 7 listed in table 4. Regarding the last question, it was not possible to link data on qualifications and experience to the survey data which covered perceptions and acceptability of the intervention as the surveys were delivered anonymously to encourage a more accurate response

In developing the methodology, the evaluation team took the research questions listed above and considered how best to collect evidence in relation to each. The process evaluation took place over the full intervention year, and evidence was collected using the methods described below.

Tracking data was collected by the project delivery team, partly through online monitoring systems and partly through their support roles with settings. The project delivery team kept records in relation to the completion, partial completion or non-completion of all of the core and optional components outlined in table 3 and table 14. These data formed evidence of the extent to which intervention group settings adhered to the intentions of the intervention, and their engagement with the programme.

The evaluation team conducted **observations of the Maths Champions Induction Sessions** for nurseries, which were led by the project delivery team. The purpose of these observations was to ensure that the evaluation considered the full Maths Champion journey for a selection of settings, from first contact with the project delivery team. The observations also allowed the evaluation team to monitor consistency in delivery between the sessions, quality of presentation and engagement of attendees. This information was used to inform the 'perceptions of the intervention' and 'outcomes of the intervention' sections of the process evaluation.

Within each setting, **survey responses from both the 'Maths Champion' and at least two other Practitioners were collected at the end of the intervention**; aiming for a total response of 45 Maths Champion surveys and 90 Practitioner surveys (Appendices R and S). The purposes of these surveys were to gather information on how the settings in the intervention group implemented the programme, their perceptions of the programme, perceived impact of the intervention, and any barriers to success.

Interviews with the project delivery team, NDNA, took place at the beginning and end of the intervention. These were conducted by the evaluation team and focused on how the programme was implemented and the relation to their original intentions, perceptions of the trial's impacts and outcomes, and any barriers (Appendices U and W).

Case studies were conducted in 10% of intervention nurseries (n=6). These were grouped into longitudinal, cross-sectional and best-practice case studies; these are listed below, along with a description of the setting. The longitudinal nurseries and cross-sectional nurseries were chosen at random by inputting the intervention nursery IDs into a random generator. From the initial four nurseries that were selected, two chose not to participate; therefore, two more nurseries were chosen at random. The best practice nurseries were selected by the NDNA, who viewed their delivery of the Maths Champion programme as successful. In each of the case study nurseries a total of three interviews were carried out: with the setting manager, a practitioner, and the Maths Champion, although in some instances the Maths Champion and the manager were completing the same role, and so were interviewed once. Consent was taken from each of these individuals (Appendix T) and interviews were recorded for the purposes of transcription. Interviews lasted around 30-90 minutes and followed their relevant interview schedule (Appendices U, V, W and X). Each interview was analysed using thematic analysis; key perceptions of the programme and perceived outcomes were pulled out and are detailed in this report.

Some information is included regarding settings' Lower-layer Super Output Area rank on the Index of Multiple Deprivation (2015); however, we can postulate that the immediate local community is rarely the main source of a setting's attendees. Nurseries attract diverse attendees due to their location, which may be near a parent/carer's place of work rather than their home. Therefore, we have relied largely upon individuals' self-reporting on the community from which the attendees are drawn, and any relevant trends in deprivation or education.

Longitudinal

Longitudinal case studies were conducted at two randomly selected nurseries, and consisted of three visits spread across the intervention period (autumn 2016, early spring 2017, and late summer 2017).

- Setting A: This was a small nursery, first registered in 2004. It was situated within a community ranked within the 50% least deprived communities using the 2015 Index of Multiple Deprivation (IMD). Its last Ofsted inspection was in 2015, where it was awarded a '2' (Good).
- Setting B: This was a large nursery, first registered in 2012. It was situated within a community ranked within the 10% least deprived communities using the 2015 IMD. Its last Ofsted inspection was in early 2017, where it was awarded a '2' (Good).

Cross-sectional

Cross-sectional case studies were conducted at a further two randomly selected nurseries, and took place at the mid-point of the intervention in early spring 2016 to allow nurseries to reach a point where action plans were developed and the majority of training to have taken place, allowing richer discussion of the programme during interviews and an accurate picture of the extent to which the programme was becoming embedded in nursery practice.

- Setting C: This was a very large nursery, first registered in 2014. It was situated within a community ranked within the 10% most deprived communities using the 2015 IMD. Its last Ofsted inspection was in 2016, where it was awarded a '2' (Good).
- Setting D: This was a medium-sized nursery, first registered in 1999. It was situated within a community ranked within the 50% least deprived communities using the 2015 IMD. Its last Ofsted inspection was in 2015, where it was awarded a '2' (Good).

Best Practice

Best practice case studies were conducted at two nurseries who were selected by the NDNA as settings which had exemplified good all-round engagement and adherence with the programme, and with whom NDNA had observed practice which was felt could be helpful in informing the process evaluation in the successes and limitations of the programme. NDNA identified these best practice case studies in June 2017.

- Setting E: This was a medium-sized nursery, first registered in 1994. It was situated within a community ranked within the 30% least deprived communities using the 2015 IMD. Its last Ofsted inspection was in 2016, where it was awarded a '1' (Outstanding).
- Setting F: This was a medium-sized nursery, first registered in 2004. It was situated within a community ranked within the 50% least deprived communities using the 2015 IMD. Its last Ofsted inspection was in 2014, where it was awarded a '2' (Good).

Costs

Data on intervention costs and time costs were collected from NDNA during interviews, other face-to-face/phone meetings and email communication. Information on intervention and time costs was also collected from participating nurseries during interviews and through the Maths Champion and Practitioner post-intervention surveys.

The cost per pupil has been calculated by dividing the cost per nursery by the average number of children eligible for inclusion in the evaluation (the children who were the target of the intervention) in nurseries participating in the evaluation. It should be noted however, that since the intervention is a whole nursery approach it could be influencing all children attending a nursery setting.

Timeline

Table 5: Timeline

DATE	TEAM	ACTIVITY
01/10/2015	ALL	Project Start
10/15 to 02/16	ALL	Protocol Development and Ethics
10/15 to 06/16	NDNA, University of Oxford	Intervention Development
01/16 to 04/16	Evaluation team	Piloting Evaluation Procedures
01/16 to 05/16	NDNA (Supported by the evaluation team)	Recruit Nursery Settings
06/16	Evaluation team (Durham)	Baseline Data Collection
06/16	Evaluation team (York)	Randomisation
07/16	NDNA	Intervention Begins
09/16	Evaluation team	Case Studies (in-depth, first visit)
01-02/16	Evaluation team	Case Studies (in-depth, second visit; cross-sectional, only visit)
05/16	Evaluation team	Case Studies (in-depth, final visit; cross-sectional, only visit)
06-07/17	Evaluation team (subcontractors A+ Education Ltd,)	Primary and Secondary Outcomes Collected
07/17	NDNA	Intervention Ends
08-09/17	Evaluation team	Parental/Guardian Follow-up on Child School Destination
09/17 to 12/17	Evaluation team	Analysis and Report Writing
01/18	Evaluation team	Submit Draft Report
02/18	Evaluation team	Respond to reviewers comments
01/18 to 03/18	Evaluation team	Pupil Matching Reference (PMR)numbers for Long Term Follow-up collected from NPD
03/18	ALL	Project End Date

Impact evaluation

Participant flow including loses and exclusions

Nursery Recruitment and Attrition

Over 300 nurseries responded to recruitment invitations by completing an online expression of interest. The nursery eligibility criteria were applied and 147 settings initially appeared to meet the eligibility criteria and were contacted by NDNA to complete and return the Memorandum of Understanding (MoU) if they wanted to take part.

125 settings agreed to participate by returning a MoU (22/147 (15%) did not return a MoU).

During the pupil recruitment phase and before nursery randomisation, 10 nurseries were excluded for the following reasons:

- **Excluded *n*=2** [nursery closed for refurbishment, *n*=1; unable to contact, *n*=1]
- **Identified as not meeting inclusion criteria *n*=8** [school/state nursery, *n*=2; too few/no children who meet eligibility criteria, *n*=5; no graduate within setting, *n*=1]

A further seven nurseries withdrew for the following reasons:

- **Other *n*=7** [evaluation too labour intensive, *n*=2; staff shortages, *n*=4; advised to withdraw by Early Years Consultant, *n*=1]

108 nurseries were randomised (Intervention *n*=54; Control *n*=54)

During the course of the evaluation, 14 nurseries (Intervention *n*=10; Control *n*=4) withdrew from the trial – 11 before pre-testing and being informed of their random allocation, and three between pre- and post-testing for the following reasons:

Prior to pre-testing, not notified of group allocation (Intervention *n*=7; Control *n*=4)

- No children meeting eligibility criteria, *n*=4 (in 3 of these cases local schools opened free provision and so children left the nursery prior to baseline data collection)
- Unresponsive to all contact, *n*=3
- Staff shortages, unable to dedicate time to project, *n*=2
- Unhappy about concept of randomisation, *n*=1
- No reason given, *n*=1

Prior to post-test (Intervention *n*=3; Control *n*=0)

- Staff shortages, unable to dedicate time to project, *n*=1
- Unresponsive to all contact, *n*=1
- All children taking part in the evaluation had left the nursery, n=1

At the time of post testing 94 nurseries remained in the evaluation (Intervention *n*=44; Control *n*=50).

Four intervention nurseries withdrew from the Maths Champions programme (but were retained for evaluation) for the following reasons:

- Lack of staff capacity, *n*=2
- Staff morale negatively affected by programme (BKSB), *n*=2

Pupil Recruitment and Attrition

Nurseries were contacted on receipt of their MoU and requested to provide information on the number of children within their setting who would be eligible to participate in the pre- and post-intervention ASPECTS assessments, i.e. at least three years of age, due to attend school in September 2017, and attending the setting for a minimum of 15 hours per week. **The total number of children across participating settings identified as eligible was 2792.**

Out of the 125 nurseries that returned a MoU, 117 (94%) provided data on pupil eligibility. The research administrator was in frequent contact with nurseries during the pupil recruitment process and parents/carers of eligible children were approached in each nursery for consent. Practitioner reports estimated that 1141 consent forms could be returned to the evaluation team.

CONSENT FORMS RETURNED TO EVALUATION TEAM: ***n*=1035**

- **TOTAL EXCLUDED/WITHDREW: *n*=10**
    - Excluded - consent form returned but incomplete, *n*=7
    - Other - parental consent withdrawn at baseline, *n*=3

TOTAL NUMBER OF ELIGIBLE AND CONSENTING CHILDREN PRIOR TO PRE-TEST: *n*=1025

- **EXCLUSION/DROP-OUT PRIOR TO PRE-TEST: *n*=44**
    - Nursery withdrew from research prior to pre- testing, *n*=10
    - Child left nursery, *n*=16
    - Child ineligible at point of testing (i.e. nursery realised the child was ineligible after parental consent had been sought), *n*=18 (too old, too young, nursery attendance <15hrs)

NUMBER OF PRE-TESTS TO BE CONDUCTED: *n*=981 (Intervention *n*=474; Control *n*=507)

NUMBER OF PRE-TESTS COMPLETED: *n*=845/981 (86.1%) (Intervention *n*=407/474 (85.9%); Control *n*=438/507 (86.4%))

Five nurseries reported that they did not have staff capacity to assess the children pre-intervention using ASPECTS and agreed to a visit by a trained member of the evaluation team to assess the children. Additional parental consent was obtained for this to happen. In total, 28 (3.3%) children were assessed by a member of the evaluation team pre-intervention. The remaining 817 were assessed by a practitioner within the setting, as per protocol. Nurseries were not informed of their allocation until they had completed the baseline testing.

- **MISSING PRE-TEST DATA: *n*=136/981 (13.9%) (Intervention *n*=67/474 (14.1%); Control *n*=69/507 (13.6%))**
    - Optional assessment (i.e. children not randomly selected for 'compulsory' testing and it was 'optional' for nursery to assess if they wanted to/had time/capacity): Intervention *n*=31; Control *n*=38
    - Absent on day of testing: Intervention *n*=1; Control *n*=1
    - Staff did not want pupil to be assessed: Intervention *n*=1; Control *n*=0
    - Unknown: Intervention *n*=34; Control *n*=30

- **EXCLUSION/DROP-OUT PRIOR TO POST-TEST: *n*=210/981 (21.4%) (Intervention *n*=131/474 (27.6%); Control *n*=79/507 (15.6%))**
    - Child left nursery: Intervention *n*=108; Control *n*=79
    - Nursery withdrew from trial: Intervention *n*=18; Control *n*=0
    - Parent withdrew consent after baseline testing: Intervention *n*=1; Control *n*=0

- Consent forms not returned to evaluation team until after post-testing period i.e. pupils were consenting and assessed at baseline by nursery practitioners, but consent forms not received by evaluation team until after post-intervention testing ceased (they'd been lost in the post). Post-testing was not arranged as evaluation team could not confirm consent at that time; however, as we had the baseline data and the consent forms at the time of analysis, baseline data is included in analysis: Intervention *n*=4; Control *n*=0

The research administrator contacted all participating nurseries via mail or telephone to collect data on child ethnicity, EYPP and children's attendance patterns prior to arranging post-test visits. Data on the number of children (with consent) who had left the nursery was collected at this point.

NUMBER OF POST-TESTS TO BE CONDUCTED: *n*=771/981 (Intervention *n*=343/474 (72.4%); Control *n*=428/507 (84.4%))

Post-intervention test assessments were conducted by 11 trained research assistants (RAs), blinded to random allocation. Two visits were arranged to each nursery on the days/times when the most children (with consent) would be present at the nursery to ensure as many children were assessed as possible. Forty-nine nurseries received a third 'mop-up' visit to collect data on, usually, one or two children. At the beginning of the post-test data collection phase, each RA received a quality assurance visit by a member of the evaluation team to ensure the assessments were being conducted as intended. No issues were highlighted. Any outstanding ethnicity and EYPP data were collected from nurseries by the RAs during the post-test visits.

NUMBER OF POST-TESTS COMPLETED: *n*=696/981 (Intervention *n*=304/474 (64.1%); Control *n*=392/507 (77.3%))

- **MISSING POST-TEST DATA: *n*=75/981 (Intervention *n*=39/474 (8.2%); Control *n*=36/507 (7.1%))**
  - Absent on day(s) of assessment: Intervention *n*=28; Control *n*=16
  - Pupil refused: Intervention *n*=2; Control *n*=2
  - Staff refused: Intervention *n*=2; Control *n*=0
  - Unknown: Intervention *n*=5; Control *n*=2
  - Child assessed but data not returned to CEM due to assessment package technical error or in a minority of cases a user saving error: Intervention *n*=2; Control *n*=16

Some children were absent on the days/times RAs were scheduled to visit the nurseries and conduct the post-tests. In order that data be captured, practitioners were requested to complete the ASPECTS post-test with these children. In total, four children (0.6%) across two nurseries were assessed at post-test by a practitioner. The remaining 692 were assessed by a RA, as per protocol.

SUMMARY OF OPT-IN CONSENT & TEST DATA

ASSESSMENTS COMPLETED AT PRE-TEST (with consent): *n*=845

ASSESSMENTS COMPLETED AT POST-TEST (with consent): *n*=696

MATCHED PRE- AND POST-TEST DATA (with consent): *n*=628

PRE-TEST DATA ONLY (with consent): *n*=217

POST-TEST DATA ONLY (with consent): *n*=68

A total of 108 nurseries were randomised, 54 to each of the interventon and control groups. Figure 2 shows a flow diagram of nurseries and participants through the trial. A total of 628 children of the 981 enrolled were included in the primary analysis; this equates to an attrition rate of 36%. It should be

noted however that it was only ‘optional’ (given capacity) to test 69 of the children missing from the analysis (Intervention *n*=31; Control *n*=38). If we discount these from the denominator, the attrition rate is estimated at 31% (1-(628/912)).

Figure 2: Nursery and participant flow diagram

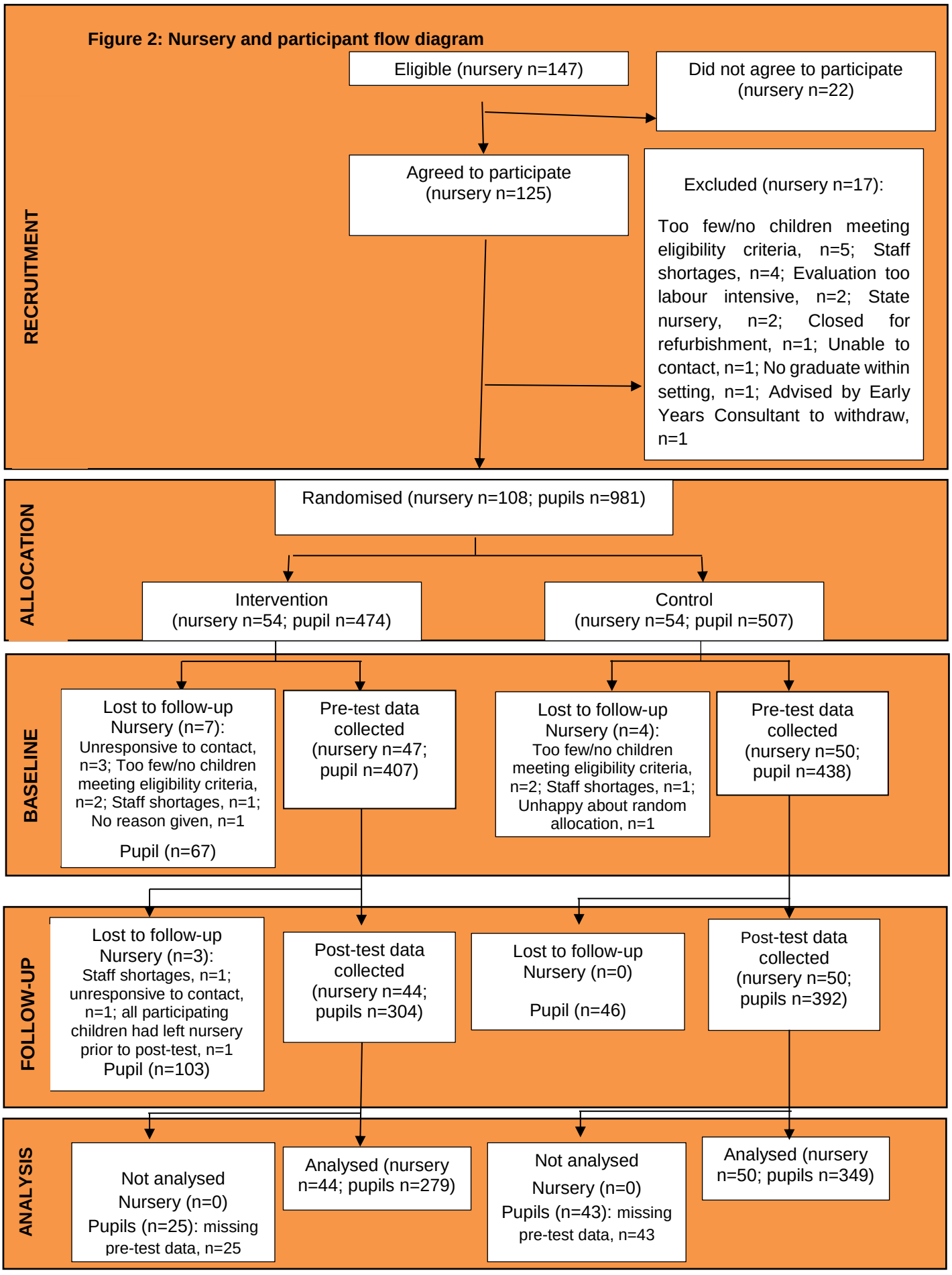


Table 6

Stage	N nurseries (n=intervention; n=control) N children (n=intervention; n=control)	Correlation between pre-test & post-test	ICC	Power	Alpha	Minimum detectable effect size (MDES)
Protocol[a]	120 (60; 60) 1,200 (600; 600)	0.70[b]	0.19[b]	80%	0.05	0.20
Randomisation[a]	108 (54; 54) 981 (474; 507)	0.70[b]	0.19[b]	80%	0.05	0.22
Primary analysis	94 (44; 50) 628 (279; 349)	0.59[c]	0.17[c]	80%	0.05	0.26

[a] allows for 10% pupil level attrition; [b] based on assumptions; [c] actual observed correlation and ICC

As designed and stated in the protocol, we aimed to have 80% power to detect an effect size of 0.2 of a standard deviation with 1200 pupils, assuming a pre-post test correlation of 0.7, an ICC of 0.19 and 10% loss-to-follow-up (Table 6). At randomisation, we anticipated to have pre-test data for 981 pupils and so, under otherwise identical assumptions, calculated the minimum detectable effect size (MDES) to be approximately 0.22. The actual observed ICC at the nursery level obtained from the primary analysis model was 0.17 (95% CI 0.10 to 0.26). The overall correlation between the pre and post-test scores of the participants included in the primary analysis was 0.59. Based on the number of pupils included in the primary analysis model (n=628), and the observed ICC and pre-post test correlation, the estimated MDES for the primary outcome was 0.26 (Table 6).

Pupil and nursery characteristics

Characteristics for the 108 randomised nurseries and 981 participating children are presented in Table 7, and appear broadly similar between the two groups. The number of children from each nursery ranged from 3 to 17 (median 10) in the intervention group, and 2 to 24 (median 10) in the control group.

Nurseries in the intervention group had, on average, 2.8 graduate staff members (2.6 in the control group). Data for the proportion of paid staff qualified to Level 3 or above is only available for nurseries who were assessed for ECERS at the end of the intervention as this was the only time that this information was collected; therefore, there is missing data for those that were not assessed for the ECERS (see section below: *Secondary outcome analyses: ECERS assessment*). On average, 78% of staff members at each nursery with this data in the intervention group were qualified to Level 3 or above (83% in the control group).

Of the 108 randomised nurseries, 65 (60.2%) were in an area falling within an Income Deprivation Affecting Children Index (IDACI) Decile of 5 or less (i.e. in the top 50% most deprived areas as determined by the IDACI): 34/54 (63.0%) in the Intervention group; 31/54 (57.4%) in the Control group.

There is a significant amount of missing data for gender, early years pupil premium (EYPP), and ethnicity. Reasons for this are as follows. Gender was intended to be collected during the pre-test ASPECTS assessment; however, the field for gender was not compulsory in the ASPECTS computer programme and nursery staff left this blank in some cases. Every effort was made to capture missing gender data at the post-testing stage, but where children had left the nurseries by this point it was not

always possible to record their gender. Ethnicity and EYPP were only collected at post-testing so these details are not available for children who were assessed at baseline but had left the nursery by the time the post-test assessments were conducted. In addition, in some nurseries, staff were not willing to provide ethnicity, gender or EYPP information for the children, particularly if they had left the nursery, and some parents asked for gender and ethnicity not to be recorded (via consent form).

For gender, there are 92 (19.4%) missing data points in the intervention group and 109 (21.5%) in the control group. For pupils with data on gender, 177/382 (46.3%) in the intervention group were male and 195/398 (49.0%) in the control group.

In relation to EYPP, 173 (36.5%) pupils in the intervention group are missing this and 98 (19.3%) in the control group. For pupils with data, 36/301 (12.0%) in the intervention group were reported to receive EYPP, and 53/409 (13.0%) in the control group.

Data on ethnicity are missing for 146 (30.8%) pupils in the intervention group and 84 (16.6%) in the control group. The vast majority of pupils, for whom ethnicity data was recorded, were white (88.7% in the intervention group and 84.2% in the control group).

Pre-test ASPECTS data was collected from 845 children across 97 nurseries (average 8.7 per nursery, SD 2.7, range 2 to 13). The average (SD) score for the mathematics component was 11.6 (6.7) out of a possible 29 in the intervention group and 11.0 (6.4) in the control group (Hedges' g effect size between the groups 0.09, 95% CI -0.04 to 0.23). The average (SD) score for the language component was 23.6 (8.7) out of a possible 53 in the intervention group and 23.1 (8.2) in the control group (Hedges' g effect size between the groups 0.06, 95% CI -0.08 to 0.19).

Children were, on average, aged 41.5 (SD 4.0) months at pre-test in the intervention group and 41.5 (SD 3.8) in the control group. Some children tested at baseline were aged less than 36 months (Intervention group 34/407 (8.4%); Control group 40/438 (9.1%)); however, all were aged over 30 months and turned three during the evaluation.

Participating children attended nursery for an average (SD) of 23.9 (11.3) hours in the intervention group and 23.4 (11.1) in the control group. There were a number of children for whom pre-test data were collected that were reported as attending nursery for less than the required 15 hours a week (Intervention group 19/407 (4.7%); Control group 27/438 (6.2%)). One explanation for this is that this data was collected only at the post-test (which also contributes to the level of missing data) at which point children may have been attending the nursery for more or less hours than they were at the start of the evaluation.

Table 7: Comparison of participating nurseries and pupils, as randomised

Variable	Intervention group		Control group	
Nursery-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
>1 graduate staff	40/54 (0)	74.1%	39/54 (0)	72.2%
≥22 children leaving for primary school in Sept 2017	27/54 (0)	50.0%	28/54 (0)	51.8%
IDACI Decile	54/54 (0)		54/54 (0)	
1	13	24.1%	13	24.1%
2	6	11.1 %	5	9.3%

3	6	11.1%	6	11.1%
4	4	7.4%	6	11.1%
5	5	9.3%	1	1.9%
6	6	11.1%	8	14.8%
7	2	3.7%	6	11.1%
8	4	7.4%	2	3.7%
9	2	3.7%	3	5.6%
10	6	11.1%	4	7.4%
Nursery-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**
N of graduate staff	54 (0)	2.8 (1.9)	54 (0)	2.6 (1.6)
N of children leaving for primary school in Sept 2017	54 (0)	23.4 (14.7)	54 (0)	24.6 (17.7)
Proportion of paid staff qualified to Level 3 or above	43 (11)	0.78 (0.16)	46 (9)	0.83 (0.13)
IDACI score	54 (0)	0.25 (0.18)	54 (0)	0.25 (0.17)
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for EYPP	36/474 (173)	7.6%	53/507 (98)	10.5%
Sex, Male	177/474 (92)	37.3%	195/507 (109)	38.5%
Ethnicity	328/474 (146)		423/507 (84)	
White	291	88.7%	356	84.2%
Mixed/multiple ethnic groups	6	1.8%	18	4.3%
Asian/Asian British	10	3.1%	31	7.3%
Black/Black British	21	6.4%	13	3.1%
Other	0	0.0%	5	1.2%
Pupil-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**
Pre-test ASPECTS mathematics	407 (67)	11.6 (6.7)	438 (69)	11.0 (6.4)
Pre-test ASPECTS language	407 (67)	23.6 (8.7)	438 (69)	23.1 (8.2)
Age at pre-test, months	407 (67)	41.5 (4.0)	438 (69)	41.5 (3.8)
Hours attendance at nursery per week	141 (333)	23.9 (11.3)	180 (327)	23.4 (11.1)

Outcomes and analysis

Primary intention-to-treat (ITT) analysis

The mean (SD) pre-test ASPECTS mathematics score, out of a possible 29, was 11.6 (6.7) in the intervention group (n=407 pupils), and 11.0 (6.4) in the control group (n=438) (Figure 3). A total of 696 (Intervention group n=304; Control group 392) pupils completed the post-test ASPECTS assessment. At post-test, the means (SD) in the two groups were 18.5 (6.2) and 17.3 (6.2), respectively (Table 9; Figure 3). Pre- and post-test ASPECTS data were both available for 628 pupils (Intervention group n=279; Control group n=349), and these were included in the primary analysis model. The characteristics of these pupils are presented in Table 8. The number of children included in the primary analysis model from each nursery ranged from 3 to 13 (median 5.5) in the intervention group, and 1 to 13 (median 7) in the control group. The mean (SD) pre-test ASPECTS mathematics score for this subset of participants was 12.6 (6.5) in the intervention group, and 11.0 (6.3) in the control group (Hedges' g effect size between the groups 0.26, 95% CI 0.10 to 0.42). The mean (SD) post-test ASPECTS mathematics score for this subset of participants was 18.7 (6.3) in the intervention group, and 17.3 (6.2) in the control group.

Figure 3: Histogram of pre- and post-intervention ASPECTS mathematics scores, for all the sample

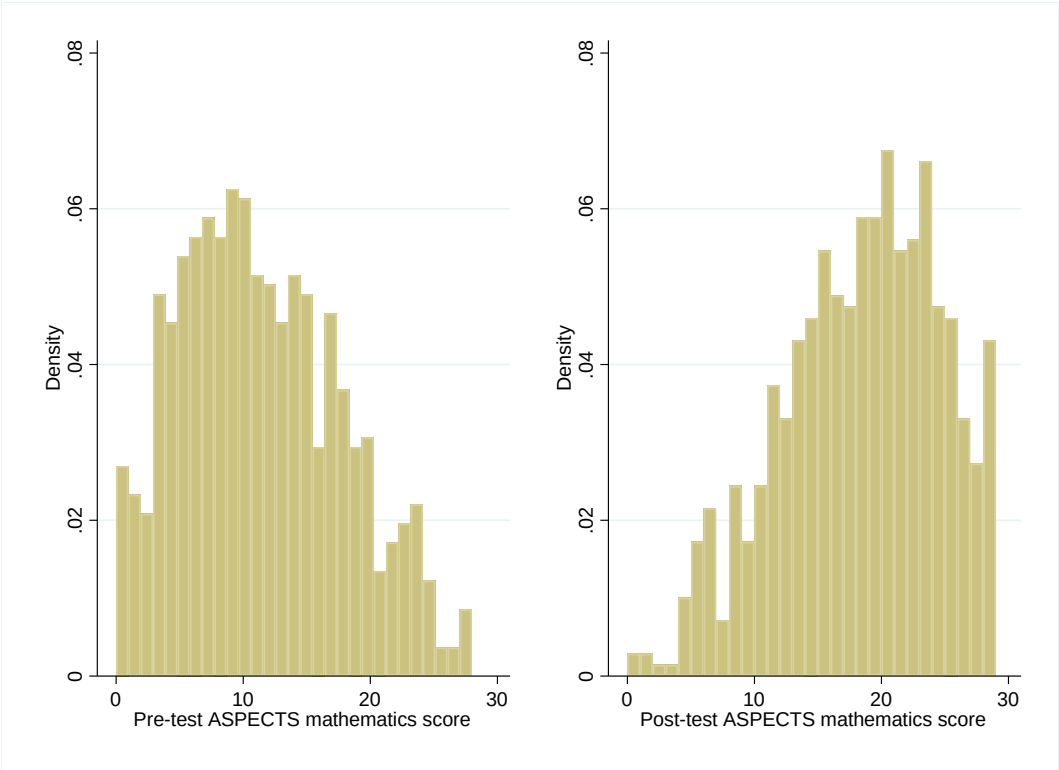


Table 8: Comparison for participating pupils, as included in the primary analysis

Variable	Intervention group		Control group	
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for EYPP	24/279 (34)	8.6%	45/349 (17)	12.9%
Sex, Male	115/279 (20)	41.2%	156/349 (35)	44.7%
Ethnicity	269/279 (10)		344/349 (5)	
White	239	88.9%	288	83.7%
Mixed/multiple ethnic groups	5	1.9%	12	3.5%
Asian/Asian British	9	3.4%	29	8.4%
Black/Black British	16	6.4%	11	3.2%
Other	0	0.0%	4	1.2%
Pupil-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**
Pre-test ASPECTS mathematics	279 (0)	12.6 (6.5)	349 (0)	11.0 (6.3)
Pre-test ASPECTS language	279 (0)	24.3 (8.3)	349 (0)	23.4 (8.2)
Age at pre-test, months	279 (0)	41.5 (3.7)	349 (0)	41.3 (3.7)
Hours attendance at nursery per week	112 (333)	25.1 (11.4)	154 (327)	23.7 (11.3)

There was no statistically significant evidence of a difference in total ASPECTS mathematics score between the intervention and control group children (mean difference 0.48, 95% CI -0.65 to 1.60, p=0.41; Table 9). The estimated effect size for the intervention effect is small at 0.10 of a standard deviation (95% CI -0.13 to 0.33) (Hedges, 2007), the equivalent of two months' additional progress. The total variance used to calculate the effect size was 24.8; the sum of 20.6 (random variation between pupils, within-cluster variance) and 4.2 (heterogeneity between nurseries, between-cluster variance).

The observed ICC at the nursery level obtained from the primary analysis model was 0.17 (95% CI 0.10 to 0.26; Table 6). The overall correlation between the pre- and post-test scores of the participants included in the primary analysis was 0.59. Based on the number of pupils included in the primary analysis model, and the observed ICC and the correlation between the pre- and post-test ASPECTS mathematics scores, the estimated minimum detectable effect size for this analysis was 0.26. The trial was initially powered to detect a 0.2 effect size, as this was considered a meaningful difference. The trial was therefore ultimately underpowered to have been able to indicate that a 0.2 effect size was statistically significant if an effect this size had been observed.

Secondary ITT analysis

The intervention effect was little changed when the model was adjusted only for baseline ASPECTS score and not the minimisation factors (mean difference 0.46, 95% CI -0.70 to 1.62, p=0.43; effect size 0.09, 95% CI -0.14 to 0.33; Table 9). The total variance was 25.4, calculated as the sum of 20.6 (within-cluster variance) and 4.8 (between-cluster variance).

Missing data

Overall, 136/981 (13.9%) pupils are excluded from the primary analysis due to missing pre-test data (Intervention n=67/474 (14.1%); Control n=69/507 (13.6%)). Reasons for this are predominantly either unknown, or because it was optional for the child to be assessed if the nursery wanted to/had time/capacity. These data are unlikely to be missing completely at random (MCAR), but it is difficult to proffer further explanation on the reasons for missingness given the limited data available for these pupils. A further 217/981 (22.1%) pupils are excluded since they are missing post-test data (Intervention n=128/474 (27.0%); Control n=89/507 (17.6%)). These were predominantly because the child had left the nursery or was absent on the day of testing. Missing pre- and post-test ASPECTS data were imputed using multiple imputation. The primary analysis model was rerun on the multiply imputed data set and Rubin's rules used to combine the treatment estimates. There was no evidence of a difference in total ASPECTS mathematics score between the intervention and control group children (mean difference 0.42, 95% CI -0.53 to 1.38, p=0.38; effect size 0.09, 95% CI -0.06 to 0.23; Table 9) following multiple imputation. The total variance was 24.5, calculated as the sum of 22.6 (within-cluster variance) and 1.9 (between-cluster variance).

Intervention fidelity and CACE analysis

All but eight intervention nurseries were assessed by the NDNA for their fidelity to the intervention. Of the eight nurseries that were not assessed, two had withdrawn from the intervention and six had withdrawn completely from the trial. These were all given an assumed fidelity score of 0/16. All control nurseries were also given a fidelity score of 0/16. RAG ratings for the eight core intervention components are summarised in Table 10. For six of the eight components, over half of the nurseries were scored green ("very engaged").

Of the 54 intervention nurseries, 44 (81.5%) were defined as engaging at least minimally with the intervention. To account for potential contamination, the control nursery that amalgamated with an intervention nursery was also assumed to have minimally engaged with the intervention. The CACE estimate of the effect of engaging at least minimally with the intervention on the pupils' mathematics attainment was a predicted increase of 0.48 points (95% CI -0.75 to 1.71, p=0.45; effect size 0.10, 95% CI -0.15 to 0.34; Table 9). Of the 54 intervention nurseries, 18 (33.3%) were defined as being very or partially engaged in *all* of the core aspects of the intervention. The CACE estimate of the effect of being very or partially engaged on the pupils' mathematics attainment was a predicted increase of 1.09 points (95% CI -1.72 to 3.91, p=0.45; effect size 0.22, 95% CI -0.34 to 0.78; Table 9).

Table 9: Primary, secondary and sensitivity ASPECTS analysis

	Raw means at post-intervention				Effect size		
	Intervention group		Control group				
Outcome: Post-test ASPECTS mathematics	n (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges' g (95% CI)	p-value
Primary ITT analysis	304 (170)	18.5 (17.8, 19.2)	392 (115)	17.3 (16.7, 18.0)	628 (279; 349)	0.10 (-0.13, 0.33)	0.41
Secondary ITT analysis	-	-	-	-	628 (279; 349)	0.09 (-0.14, 0.33)	0.43
Results following MI	-	-	-	-	981 (474; 507)	0.09 (-0.06, 0.23)	0.38
CACE: any vs. no engagement	-	-	-	-	628 (279; 349)	0.10 (-0.15, 0.34)	0.45
CACE: good vs unsatisfactory engagement	-	-	-	-	628 (279; 349)	0.22 (-0.34, 0.78)	0.45

Table 10: Summary of fidelity RAG rating for core components of the Maths Champions intervention as assessed by NDNA

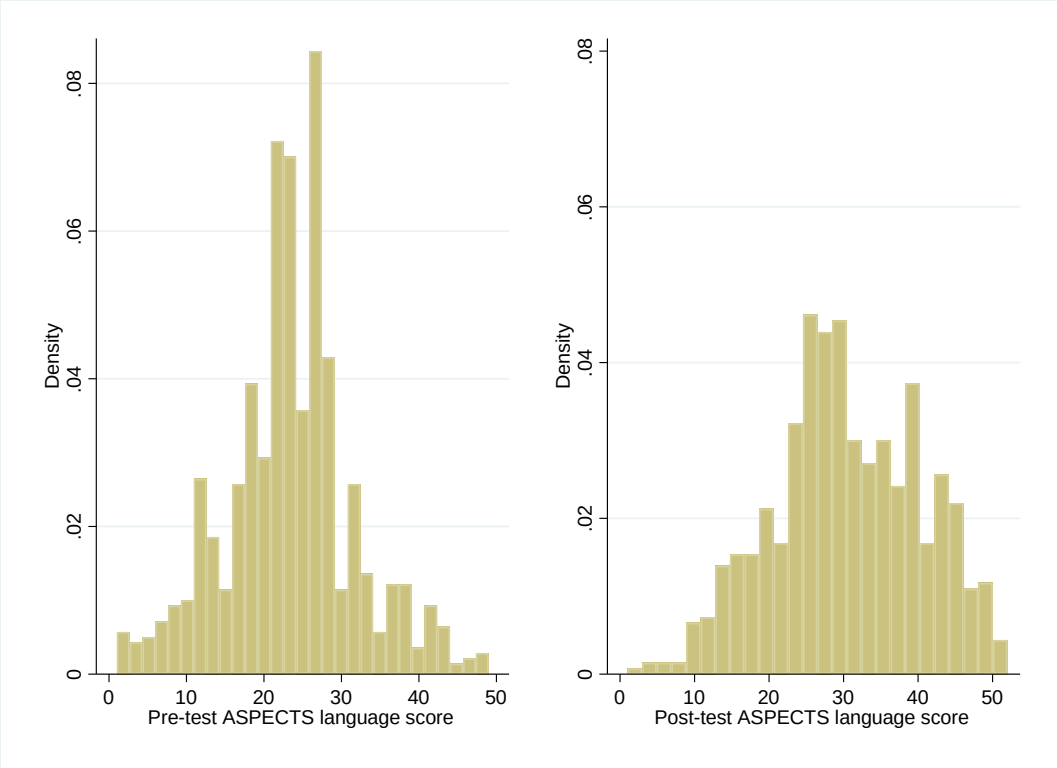
Core Components	RAG rating	Number (%) of intervention nurseries (n=54)
Attendance by the Maths Champion at the seven online training courses	Green = 2	43 (79.6)
	Amber = 1	0 (0.0)
	Red = 0	11 (20.4)
Attendance by the Practitioner at five of these online training courses	Green = 2	27 (50.0)
	Amber = 1	8 (14.8)
	Red = 0	19 (35.2)
Attendance at live webinar training sessions	Green = 2	34 (63.0)
	Amber = 1	0 (0.0)
	Red = 0	20 (37.0)
Attendance at recorded webinar training sessions	Green = 2	19 (35.2)
	Amber = 1	6 (11.1)
	Red = 0	29 (53.7)
Completion of the 11 core activities	Green = 2	25 (46.3)
	Amber = 1	6 (11.1)
	Red = 0	23 (42.6)
Completion of BKSB and follow-up activities	Green = 2	28 (51.8)
	Amber = 1	0 (0.0)
	Red = 0	26 (48.2)
Completion and continued use of an action plan	Green = 2	31 (57.4)
	Amber = 1	6 (11.1)
	Red = 0	17 (31.5)
Completion of action research throughout the project	Green = 2	31 (57.4)
	Amber = 1	0 (0.0)
	Red = 0	23 (42.6)

Secondary outcome analyses

ASPECTS language score

The mean (SD) pre-test ASPECTS language score, out of a possible 53, was 23.6 (8.7) in the intervention group (n=407), and 23.1 (8.2) in the control group (n=438). (Table 5; Figure 4). At post-test, the means (SD) were 31.6 (10.3) and 29.7 (9.8), respectively (Table 11; Figure 4).

Figure 4: Histogram of pre- and post-intervention ASPECTS language scores, for all the sample



The mean (SD) pre-test ASPECTS language score for pupils with both pre- and post-test scores, out of a possible 53, was 24.3 (8.3) in the intervention group, and 23.4 (8.2) in the control group (Hedges' g effect size between the groups for this subset 0.11, 95% CI -0.05 to 0.27). The mean (SD) post-test ASPECTS language score for this subset of participants was 31.8 (10.3) in the intervention group, and 29.6 (9.9) in the control group.

There was no statistically significant evidence of a difference in total ASPECTS language score between the intervention and control group children (mean difference 1.41, 95% CI -0.53 to 3.35, p=0.15; Table 11). The estimated effect size for the intervention effect is 0.17 of a standard deviation (95% CI -0.06 to 0.40) the equivalent of 2 months' additional progress. The total variance was 68.3, calculated as the sum of 56.0 (within-cluster variance) and 12.3 (between-cluster variance).

Table 11: Secondary ASPECTS analyses

Outcome	Raw means at post-intervention				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges' g (95% CI)	p-value
Post-test ASPECTS language	304 (170)	31.6 (30.4, 32.7)	392 (115)	29.7 (28.8, 30.7)	628 (279; 349)	0.17 (-0.06, 0.40)	0.15

ECERS assessment

Ninety-one (84.3%) of the 108 randomised nurseries were assessed for ECERS at the end of the evaluation (43/54 (79.6%) in the intervention group, and 48/54 (88.9%) in the control group). The mean (SD) score for the ECERS-3, out of a possible 7, was 3.1 (0.7) in the intervention group and 3.2 (0.7) in the control group. No evidence of a difference was observed between the intervention and control groups in the total ECERS-3 score (-0.08, 95% CI -0.37 to 0.20, p=0.56; effect size -0.12, 95% CI -0.54 to 0.29). In the sensitivity analysis removing the two amalgamated nurseries, the treatment effect estimate was materially unchanged (adjusted mean difference -0.08, 95% CI -0.37 to 0.21, p=0.59).

The mean (SD) score for the maths subscale of the ECERS-E, out of a possible 7, was 3.5 (1.1) in the intervention group and 3.1 (0.8) in the control group. No evidence of a difference was observed between the intervention and control groups in the maths subscale of the ECERS-E (0.36, 95% CI -0.06 to 0.77, p=0.09; effect size 0.36, 95% CI -0.06 to 0.78). [Sensitivity analysis: adjusted mean difference 0.37, 95% CI -0.05 to 0.80, p=0.09.]

The mean (SD) score for the composite maths score, out of a possible 14, was 6.3 (2.0) in the intervention group and 5.5 (1.4) in the control group. There was evidence to suggest that the composite maths score for nurseries allocated to the intervention was statistically significantly higher than the score for nurseries allocated to the control group (0.79, 95% CI 0.05 to 1.54, p=0.04; effect size 0.46, 95% CI 0.03 to 0.88). [Sensitivity analysis: adjusted mean difference 0.82, 95% CI 0.06 to 1.58, p=0.04.]

Practitioner confidence survey

A total of 272 practitioner confidence surveys (Intervention group n=118; Control group n=154) were received from 88 nurseries (41 in the Intervention group and 47 in the control group). A response from the Maths Champion of a nursery could be identified for 30 (73.2%) of the 41 Intervention nurseries. Responses to the individual items are presented in Appendix AA in Tables 1 to 6 for all respondents by randomised group and just for the Maths Champions in the Intervention group (where these could be identified).

No statistically significant evidence of a difference was observed between the intervention and control groups in the scores for subscale 1 (Beliefs about Nursery Aged Children and Maths), subscale 2 (Confidence in Helping Nursery Aged Children Learn Maths), or subscale 3 (Confidence in Own Maths Abilities) (Table 12). The adjusted mean difference between the intervention and control groups for subscale 1 was 0.39 (95% CI -0.31 to 1.09, p=0.27); the effect size was calculated as 0.16 (95% CI -0.14 to 0.47). The total variance used to calculate this effect size was 5.7; the sum of 4.7 (random variation between pupils, within-cluster variance) and 1.0 (heterogeneity between nurseries, between-cluster variance). The adjusted mean difference between the intervention and control groups for subscale 2 was 0.98 (95% CI -0.38 to 2.35, p=0.16); the effect size was calculated as 0.22 (95% CI -0.09 to 0.52). The total variance used to calculate this effect size was 20.5; the sum of 16.2 (random

variation between pupils, within-cluster variance) and 4.3 (heterogeneity between nurseries, between-cluster variance). The adjusted mean difference between the intervention and control groups for subscale 3 was 0.39 (95% CI -1.17 to 1.96, p=0.62); the effect size was calculated as 0.07 (95% CI -0.20 to 0.34). The total variance used to calculate this effect size was 33.6; the sum of 32.6 (random variation between pupils, within-cluster variance) and 1.0 (heterogeneity between nurseries, between-cluster variance).

Table 12: Practitioner Confidence Subscale Summary Results

	Raw means					Effect size	
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Beliefs About Nursery Aged Children and Maths	112 (6)	28.3 (27.9, 28.7)	143 (11)	28.0 (27.6, 28.5)	233 (99; 134)	0.16 (-0.14, 0.47)	0.27
Confidence in Helping Nursery Aged Children Learn Maths	116 (2)	47.3 (46.5, 48.2)	148 (6)	46.2 (45.3, 47.1)	243 (102; 141)	0.22 (-0.09, 0.52)	0.16
Confidence in Own Maths Abilities	113 (5)	29.6 (28.6, 30.6)	136 (18)	28.7 (27.6, 29.9)	227 (99; 128)	0.07 (-0.20, 0.34)	0.62

Subgroups: number of hours child attends nursery, EYPP and gender

Summary statistics for the ASPECTS mathematics score are presented in Table 13 by average number of hours the child attends the nursery (dichotomized at the median of 20 hours), EYPP status, and gender. These summaries indicate that, in general, children attending nursery for more than 20 hours performed slightly better on the post-test than those who attended nursery less than this. Scores for children who were eligible for the EYPP tended to be lower than for those who were not. Scores were similar for male and female pupils. In adjusted regression analyses that included interaction effects, the effect of the intervention on post-test ASPECTS mathematics score was not seen to be substantially altered by the average number of hours the child was reported to attend the nursery (as a continuous variable, interaction effect p=0.75), having EYPP status (interaction effect p=0.58), or gender (interaction effect p=0.56).

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

	Raw means			
	Intervention group		Control group	
Outcome	n	Mean (SD)	n	Mean (SD)
Number of hours at nursery:				
≤20	52	18.0 (6.1)	82	16.0 (6.6)
>20	67	20.0 (6.1)	76	17.4 (5.9)
EYPP:				
Yes	27	17.1 (7.4)	47	14.6 (6.4)
No	237	18.7 (6.1)	328	17.8 (6.0)
Gender:				
Male	117	18.8 (6.6)	156	16.8 (6.0)
Female	144	18.6 (6.2)	158	17.6 (6.3)

Cost

Cost during evaluation

The cost of delivering the intervention during the evaluation has been calculated at £350 per nursery[2] plus the additional cost of regional induction events (which includes venue hire, travel and NDNA staff delivery time) and face-to-face nursery visits conducted by NDNA (which includes travel and NDNA staff time) at a total cost of £6944.78, divided by the number of intervention nurseries (54) which results in an additional cost of £128.60 per nursery. This results in a total cost of Maths Champions per nursery during the evaluation of £478.60 for year 1.[3]

The current and predicated future commercial cost of Maths Champions to extend for one further year is an average cost of £85[4]

The total cost of Maths Champions, per nursery, over a three-year period is therefore £648.60. The total cost of Maths Champions, per nursery, per year is £216.20.

[2] The commercial cost of Maths Champions (1st year) (to non-evaluation settings) during evaluation period was £330 for NDNA members, and £370 for non-members. An average cost of £350. (http://www.ndna.org.uk/NDNA/Community/Maths_Champions.aspx).

[3] The current and predicated future cost of Maths Champions (1st year) is £400 for NDNA members and £500 for non-members (NDNA do not report any anticipated changes to the cost in immediate future years).

[4] The current and predicated future cost to extend Maths Champions for one further year is £75 for NDNA members, and £95 for non-members. An average cost of £85.

The cost of Maths Champions to nurseries during the evaluation was completely waivered. Costs were met by NDNA through funding from EEF. (Note: 31 of the 54 nurseries allocated to the intervention group were NDNA members.)

Pre-requisites

Several nurseries raised the issue of the monetary costs of taking part in the Maths Champions programme. This tended to be spent on creation of resources and printing and paying for cover so that staff could have time to complete their Maths Champions work.

Whilst implementing the Maths Champions programme nurseries needed to make use of resources and natural objects which would often be found in nursery settings e.g. mud kitchens, water play, stones etc.

Staff time

The Maths Champions intervention is a whole nursery, all staff approach, and as such is delivered, to some degree, by all staff throughout the nursery day; however, the identified Maths Champion and practitioners should spend time 'training' or in professional development. During the initial set-up of the programme (which could take around 2 months to complete and which included activities of completing staff registration forms, 2 initial online courses, evaluating settings practice using audit tools and developing an action plan), the average amount of time per week that the Maths Champions themselves spent on related professional development work was 3.7 hours; with 50% doing this in their non-working time (as reported by 'Maths Champions' in the post-intervention survey). Once the programme was established in the nursery, this workload dropped to an average of 2.9 hours per week; with 42% working in their own time (Appendix Y). This workload was less for practitioners; although during initial set-up 18% of practitioners reported that they completed their professional development work outside of their work hours, increasing to 22% after initial set-up, as reported in the post-intervention survey (Appendix Z). There is no reason to suspect that if the programme continued to be implemented (for an additional 2 years) that the average workloads reported, 2.9 hours for Maths Champions would be any different in future years.

Cost per pupil

In participating nurseries the average number of children identified by the nursery as eligible for the evaluation (that is being three years old at the start of the intervention, due to attend school in September 2017 and attending nursery for a minimum of 15 hours per week) was 24 (based on 117 nurseries providing information on the number of eligible children, a total of 2792).

The cost per pupil has been calculated based on the cost of Maths Champions during the evaluation period plus the additional costs associated with changes to the Maths Champions programme made during the evaluation period.

Table 14: Costs

Item	Type of Cost	Cost Year 1	Cost Year 2	Cost Year 3	Total cost over 3 years	Total cost per pupil per year over 3 years
Maths Champions Package 1 year (During evaluation Cost)	Start up cost per nursery	£350	£0	£0	£350	
Maths Champions Induction sessions and face-to-face support	Start up cost per nursery	£128.60	£0	£0	£128.60	
Maths Champions extend 1 year (Current Cost)	On-going running cost	£0	£85	£85	£170	
Total					**£648.60**	**648.60/3/24 = £9.01**

Process evaluation

Implementation/Perceptions of the intervention

Research questions:	Dimensions:
3. What are the different stakeholder viewpoints on the intervention?	• Participant responsiveness • Programme differentiation
4. What are the key success factors required for the Maths Champions intervention to work well?	• Quality • Monitoring of control/comparison conditions
5. What are the barriers to successful delivery of the intervention?	• Quality • Monitoring of control/comparison conditions

Here we discuss the key themes emerging from the process evaluation that relate to the setting staff's perceptions of the Maths Champions (MC) programme, and the resulting conclusions that can be drawn regarding barriers and facilitators to success. These have been categorised as training, engagement, issues surrounding time, money and staffing, resources and the Basic and Key Skill Builder assessment. Evidence is drawn from all data collection sources including interviews and surveys. A total of 5 best practice and 5 cross sectional case study interviews were conducted and 11 longitudinal case study interviews. A total of 38 Maths Champion post-intervention surveys and 76 Practitioner post-intervention surveys were received in time for analysis. For both surveys, this was a response rate of 84%.

Training

Interview data indicated that nursery managers chose to take part in the MC programme because it allowed their staff opportunities for continuing professional development (CPD), particularly in the current financial climate where many nurseries report to be struggling and thus do not have available funds to train staff.

> *'That is one of the reasons why we decided and elected to take part in Maths Champions because it was something that we were going to benefit from in terms of staff training'.* (Manager, Setting E, Best Practice)

As part of the Maths Champion and practitioner surveys, the respondents were asked to rate various aspects of the MC training (Appendices Y and Z). Generally, ratings were positive, although did differ between Maths Champions and Practitioners. For their overall rating of the online training, 32/38 (84%) of Maths Champions and 58/76 (76%) of Practitioners felt training was 'very' or 'quite' useful. When asked to rate the monthly webinars, 25/38 (66%) of Maths Champions felt that they were 'very' or 'quite' useful, compared to 33/76 (43%) of Practitioners. For a full comparison of each training course, please see Table 15.

Table 15: Practitioner vs Champion ratings of courses

Maths Champion training course	% of Practitioners who rated course as 'quite' or very' useful (number of respondents)	% of Maths Champions who rated course as 'quite' or 'very' useful (number of respondents)
Maths in the Early Years	80% (61/76)	92% (35/38)
Let's look at number	75% (57/76)	92% (35/38)
Let's look at shape, space and measure	76% (58/76)	89% (34/38)
Train the Trainer	NA	79% (30/38)
Evaluating your Practice and Impact Measurement	NA	82% (31/38)

Staff engagement

In the post-intervention survey, Maths Champions indicated high levels of engagement from nursery staff; 23/38 (61%) reported that staff were either 'quite' or 'very' engaged at the start of the MC programme. This increased to 27/38 (71%) throughout the MC programme (Appendix J). Case studies indicated that staff were generally motivated to improve maths provision in their nurseries and felt that taking part in MC allowed them to do this.

Interview responses showed that nurseries had generally positive experiences of being able to reflect on their maths practice and focus on how to improve it.

> *'It was really good to have the opportunity, I suppose, to make the time to sit back and look at what we were doing and be supportive to improve in the areas that we found we wanted to work on'.* (Manager Interview, Setting E, Best Practice)

Nonetheless, some Maths Champions reported that not all staff had high levels of engagement. The main reasons given for this were staff apprehension about workloads and assessments, low levels of maths confidence or lack of time to give the programme any real focus. Responses indicate that staff were more focussed on their practice with children as opposed to improving their own knowledge.

> *'The staff were more engaged in parts of the programme concerning children, activities and outcomes and those e-learning courses than refreshing/learning for their own CPD using BKSB'.* (Maths Champion Survey)

The project delivery team discussed in their end-point interview, the difficulty of achieving high levels of staff engagement, highlighting how important their support is in achieving this.

> *'That has been a real hurdle to get over…to get them engaged, especially when it is an online programme and you are not there in their face 24/7'* (Project delivery team end-point interview)

Time, cost and staffing

Overall, the delivery of the Maths Champions programme was reported to be very time consuming, with many staff in both case study interviews and surveys mentioning this as a significant issue.

> *'I would say [Maths Champions took] a good 20 hours a week some weeks at the beginning. I used to take things home'* (Maths Champion interview, Setting C, Cross-sectional)

> '*[the webinars] are quite good. I mean, time is always very precious and it is quite a chunk to get through and sit through but I find I do it on an evening at home'*.(Maths Champion interview, Setting D, Cross-sectional)

During the initial set-up of the programme, the average amount of time per week that the Maths Champions themselves spent on related professional development work was 3.7 hours; with 19/38 (50%) doing this in their non-working time. Once the programme was established in the nursery, this workload dropped to an average of 2.9 hours per week; with 16/38 (42%) working in their own time (Appendix Y). This workload was less for practitioners, although during initial set-up 14/76 (18%) of practitioners reported that they completed their professional development work out of their work hours, increasing to 17/76 (22%) after initial set-up (Appendix Z).

Whilst high workload was seen as an issue, some nurseries felt that this was beneficial.

> *'Obviously there is a lot for me to do, you know, pretty much everything we've done we've found had benefitted us so has been worthwhile.'* (Maths Champion interview, Setting E, Best Practice)

Case studies and survey responses also exposed differing abilities to handle this workload. Nurseries who had more success in this area reported good team collaboration and time management techniques. Setting E, which was selected as a 'Best Practice' nursery, demonstrated that a strong ethos of teamwork and collaboration not only facilitated a more productive work environment, but also increased staff enjoyment of the programme.

> *'We've got a good team here […] you get an idea and everyone adds to it and they chip in and bring the resource in […] it has been quite fun actually'.* (Practitioner interview, Setting E, Best Practice)

As the aforementioned survey statistics demonstrate, work was often completed in staff's own time. Some nurseries took the approach of paying their staff overtime for this; however, the manager of Setting B expressed a belief that staff needed to be allowed time within their work hours to complete work associated with the project, and there should not be an expectation on staff to do overtime.

> *'[Practitioners] already work long hours and overtime is not what people want. People want time to actually do their job […] I can give [the Champion] a day off upstairs in the office to do it so that she can just get it done without interruption and having her work covered downstairs'* (Manager interview, Setting B, End-point)

Although many nurseries felt that the initial workload for Maths Champions was time exhaustive, particularly the initial resource creation, audits and creation of action plans, one nursery felt that the more holistic approach of Maths Champions meant that less time was needed. This member of staff felt more able to see opportunities for maths in children's play, as opposed to planning for it in a discrete way.

> *'You used to plan for it before but I think that we are doing it in the moment because we can see what they want to achieve and what they want to learn'.* (Practitioner interview, Setting F, Best Practice)

A topic which repeatedly came up in staff interviews was the issue of releasing staff. Due to ratio requirements of nurseries, settings found it hard to release staff from their designated group of children in order to complete training and assessments. This was a particular concern for Setting A, who

expressed concerns regarding staff sickness or turnover as a result of workload and related stress. The Maths Champions programme was found to be time consuming to this nursery, and the management did not wish to push staff to complete the programme in its entirety if it was at the expense of staff welfare.

> *'In this sector [it is challenging] to retain staff, so I don't want to put them under any undue pressure that might affect that'.* (Manager interview, Setting A, Mid-point)

Several nurseries raised the issue of the monetary costs of taking part in the Maths Champions programme. This tended to be spent on creation of resources, printing and paying for cover so that staff could have time to complete their Maths Champions work. One nursery expressed their view that, although taking part in the programme did cost them financially, management chose to see this cost as paying for staff CPD.

> *'Can you get equivalent course for £10 an hour..? Probably not'.* (Manager interview, Setting B, End-point)

Resources

One particularly positive element of the project was the staff opinions of MC resources. Survey responses showed that 54/76 (71%) of practitioners and 34/76 (89%) of Maths Champions viewed the resource bank as 'quite' or 'very' useful (Appendices Y and Z). The resource audits were found to be useful for allowing staff to plan what they needed to improve in their provision. Although lack of time was an issue for many staff, the bank of resources was found to somewhat ease the planning time needed.

> *'The resources are definitely a really good idea cos (sic) it is a way for me to push the maths but without having to sit and plan everything myself cos (sic) the activity plans are already there […] I am really enjoying that'* (Maths Champion interview, Setting B, Mid-point).

Further, the majority of staff found resources to be appropriate for the children's stage of development, with 66/76 (87%) of Practitioners and 38/38 (100%) of Maths Champions rating the resources as 'mostly' or 'completely' appropriate (Appendices Y and Z). Practical resources allowed maths to be taught in a hands-on, play-based way, which was found to be a benefit in the eyes of nursery staff.

> *'They've allowed us to shake off that reputation that maths […] is something that is strictly for learning so heads down and done at a desk […] you see maths on a practical and very fun level […] it involves a lot of creative thinking and working out problems'*. (Manager interview, Setting E, Best Practice)

Basic and Key Skill Builder (BKSB) test

Although staff reported feeling that their overall confidence increased as a result of the programme (see section below on Staff Confidence), respondents to the survey and interviewees reported that the BKSB audits had the unintended consequence of negatively affecting staff confidence. In the Practitioner and Maths Champion surveys, only 50% of respondents (38/76 for Practitioners and 18/36 Maths Champions) found the BKSB assessments 'quite' or 'very' useful (Appendices Y and Z). A large number of staff questioned felt that their BKSB audits were too difficult and not necessary, due to the discrepancy between the level of maths in the tests (GCSE level maths) and the maths they were teaching the nursery children. Not being able to achieve a high mark resulted in staff feeling less confident about their abilities. Due to lack of time already being an issue, the BKSB audits were seen as a 'waste of time' when the staff could be doing activities that they felt were more important.

> *'I am not going to teach these kids Pi or algebra […] it's time consuming'* (Practitioner interview, Setting B, Initial interview)

> *'I thought it would just kill [the staff's] confidence and I thought we were trying to do the opposite, building our confidence'. 'We were setting them up to fail'.* (Maths Champion interview, Setting A, Mid-point)

The BKSB audit was discussed during an initial interview with the project delivery team, who described its importance.

> *'You are not going to get the measurement of progression of practitioners if they don't use BKSB, because that is a tool that they evaluate themselves against [...] So we won't get a good measure, we won't have the robust data that we need'.'* (Project delivery team initial interview)

However, in the end-point interview, the project delivery team discussed that this requirement may now be less important.

> *'The barrier of BKSB, which has [been] identified for quite some time, will alleviate now because we don't have to use a GCSE requirement [for practitioners'] and it now more about confidence building and getting them to functional skills level'.* (Project delivery team end-point interview).

Fidelity

Research questions:	Dimensions:
1. To what extent are the nurseries involved engaging with and delivering the intervention?	• Fidelity • Dosage • Reach • Participant responsiveness
6. Is fidelity to the intervention being maintained?	• Fidelity • Adaptation

The NDNA's regular contact with intervention settings enabled them to gather appropriate data for the monitoring and measuring of implementation fidelity of nurseries delivering the project. This data, described below, was provided to the evaluation team at intervals during the intervention, and then finalised into an agreed model of measurement by the evaluation team with the project delivery team's support and guidance.

Fidelity was broadly defined as a measure of 'delivery as intended', but was further conceptualised into two components. Firstly, *procedural* – 'the extent to which the implementer adheres to the structure and sequence of activities outlined by the intervention developer' (Humphrey et al 2016; O'Donnell, 2008). This component is drawn from objective data, and broadly addresses the fidelity, dosage and reach dimensions as defined by Humphrey et al (2016). It is constructed from a measure of each setting's engagement with the core components of the programme. The second component, *participant responsiveness*, captures some of the wider aspects of 'delivery as intended'. It is constructed from a measure of each setting's engagement with the optional components of the programme.

Procedural fidelity

Procedural fidelity was defined by the project delivery and evaluation teams jointly, and consists of eight criteria. These can be broadly split into 1) attendance data from the online training courses and live sessions, which tracks attendance per individual and per setting on each live and recorded session;

and 2) monitoring data gathered by NDNA during face-to-face, email and telephone contact with the setting over the duration of the trial.

1. Attendance
    - Attendance by the Maths Champion at the seven online training courses
    - Attendance by the Practitioner at five of these online training courses
    - Attendance at live and recorded webinar training sessions

2. Monitoring
    - Completion of the 11 core activities
    - Completion of the Basic and Key Skill Builder (BKSB) internal assessment and follow-up activities
    - Completion and continued use of an action plan
    - Completion of action research throughout the project

Performance against each of these criteria was considered and finalised by the evaluation and project delivery teams, and is defined in Table 1 (provided earlier in the report).

These data were gathered by the project delivery team during the trial and provided to the evaluation team at the close of the trial in summer 2017.

In order to aid analysis, the 'RAG' rating status was assigned a numerical value (0, 1 or 2) to create a final score per nursery out of 16. It was considered by both teams that within these criteria, no single criterion was more important than another, and that as a measure of fidelity each of the criteria should be considered equal. As such, the scale of 0-16 was used as the final scale to measure procedural fidelity for each setting. A summary of the ratings for the 46 observed nurseries is presented in Table 8 (provided earlier in the report).

Good fidelity was defined as the nursery being rated by the NDNA as being very or partially engaged in *all* of the core components of the intervention i.e. all core components rated green or amber (minimum score of 8 and all components scoring at least 1). A third (n=18) of the 54 settings randomly allocated to the intervention group achieved a good fidelity rating. This group includes both nurseries chosen by the project delivery team as representing 'best practice', selected for the case studies. Within this group all settings achieved 2 marks (very engaged) in the first criterion (attendance by the Maths Champion at the seven online training courses). This may indicate the nurseries consider this component to be integral to taking part in Maths Champions.

Partial/Unsatisfactory fidelity was defined as nurseries for whom at least one but not all core components of the intervention was rated red; 26 (48.1%) of the 54 settings randomly allocated to the intervention group obtained this rating. Within this group, most settings (n=25, 96.2%) achieved 2 marks (very engaged) against the first criterion, attendance by the Maths Champion at the seven online training courses.

No intervention implemented was defined as all core components of the intervention being rated red (score of 0); 10 (18.5%) of the 54 settings randomly allocated to the intervention group did not implement the intervention.

Participant Responsiveness/Engagement

The second component, *participant responsiveness,* captures some of the wider aspects of 'delivery as intended'. It is constructed from a measure of each setting's engagement with the four optional components of the programme, as defined by the project delivery team, giving an indication of the extent to which the nursery adopted and integrated the ethos of the intervention into their wider approach, and

were motivated to develop and progress. The project delivery team highlighted these optional components to the evaluation team during the evaluation period after the protocol for this project had been signed off. The project delivery team defined the thresholds used for these optional components.

By monitoring sign-up and usage, and through discussion with the setting's staff, the project delivery team gauged, for each setting, the following:

- Staff training activity; have staff used opportunities for peer observation, staff training sessions, mentoring, and supervision?
- Resource bank; have staff accessed and engaged with the resource bank?
- Pinterest; have staff accessed and engaged with the Maths Champions Pinterest board, and to what extent has it added depth to the programme?
- Facebook; have staff joined the Maths Champions group, are they an active member, how frequently do they use it?

Against these criteria, it was decided that performance could not be separated into three levels, and so instead the project delivery team gave either a 'green' or 'red' rating against each criterion. These are detailed in Table 15, with the number and percentage of nurseries that scored each rating out of the 46 observed.

Table 15. Optional Components Fidelity Rating

Optional Components	Description	RAG rating	Nurseries (n=46), n (%)
Staff Training Activity (peer observation, staff training, mentoring, supervision)	Used opportunities	Green	29 (63.0)
	Not used	Red	17 (37.0)
Resource bank	Activities accessed regularly and used	Green	38 (82.6)
	Not used	Red	8 (17.4)
Pinterest	Used frequently	Green	29 (63.0)
	Not used	Red	17 (37.0)
Facebook	Active member, uses a lot	Green	37 (80.4)
	Not a member	Red	9 (19.6)

Again, it was considered by both the project delivery and evaluation team that within these criteria, no one criterion was more important than another, and that as a measure of fidelity each of the criteria should be considered equal. As such, a numerical value of 0 or 1 was assigned to red and green respectively, and a total mark of 4 was attributed per setting.

Good fidelity was defined as full marks against each criterion (total of 4 marks); 21 (38.9%) of the 54 settings achieved a 'good' rating. This group includes both nurseries chosen by the developer as representing 'best practice', selected for the case studies.

Partial/Unsatisfactory fidelity defined as those nurseries achieving a mixture of red and green (0 or 1) scores, but not rated as either all red or all green for the four optional components. 21 (38.9%) of the 54 settings achieved a 'partial/unsatisfactory' rating.

No optional components engaged with was defined as all optional components of the intervention being rated red (score of 0); 12 (22.2%) of the 54 settings randomly allocated to the intervention group fell into this category.

Perceived outcomes of the intervention

Research questions:	Dimensions:
2. To what extent have nursery practitioners changed their practice from the beginning to the end of the intervention?	• Quality • Participant responsiveness
4 What are the different stakeholder viewpoints on the intervention?	• Participant responsiveness • Programme differentiation

Here we discuss the key themes emerging from the process evaluation that relate to the perceived outcomes of the Maths Champions programme. These are categorised as: staff confidence, children's progress, and parental engagement. Evidence is drawn from all data collected.

Staff confidence

The majority of practitioners who completed the survey described their role in the MC programme as a positive experience. Staff felt that the programme had helped to develop their knowledge and skills; thus improving their early years teaching practice.

> *'Maths Champions has not only helped me [with] planning activities, implementing maths various ways, it has also helped my understanding of maths better which I can pass on to other practitioners and students.'* (Practitioner survey)

Confidence was an additional factor which was heavily discussed in both interviews and the survey responses. However, opinions on this did vary widely. Overall, people believed that their maths knowledge and teaching skills had improved as a result of taking part in MC, meaning that they felt more confident about integrating maths into their practice and the setting in general.

> *'Practitioners are far more aware of how to promote maths within play (child and adult led) and the environment. Practitioners understand why early foundations in maths are important and their role in supporting this.'* (Maths Champion Survey)

> *'It has hugely helped remove the negative stigma attached to maths and allow practitioners and children to enjoy learning maths together. Confidence has grown all around nursery!'* (Practitioner survey)

Quality of Maths Provision within Nursery Setting

In settings where engagement with the Maths Champions activity was high, respondents described the quality of maths provision within the nursery setting as improving. Staff felt that this was due to higher levels of confidence in staff as a result of the programme, and breadth of experience and knowledge of appropriate activities and resources that had been improved due to the programme.

Some nursery managers identified that, as a result of the programme, maths provision was now integrated within play and the daily routine, as opposed to being separated into focused activities. This was felt to be a signifier of higher quality provision.

> *'The girls don't really need to think about planning it out as much now because they can see when maths happens naturally while the children are playing, and they make the most of those moments.'* (Practitioner survey)

> *'Staff embedded maths in children's activities and used every opportunities to encourage children to learn maths e.g. count plates or spoons at lunch time.'* (Practitioner survey)

Children's progress

Staff's increased awareness of maths and how to integrate it into their practice, as described above, was seen as having a positive effect on the children's learning.

One of the perceived benefits of MC was the increased maths abilities of the children. Moreover, children were found to be more enthusiastic about maths and showed more confidence in their abilities.

> *'The children have wanted to take more part in developing maths, for example wanting more maths activities out of the cupboard.'* (Practitioner survey)

> *'Children have been more curious, and have been asking more questions.'* (Practitioner survey)

Additionally, nursery staff have found that children have shown quantifiable improvements in their mathematic skills, as demonstrated by the nursery's own in-house monitoring.

> *'The cohort this year are far more interested in maths activities and a higher number of them are meeting higher levels of development in terms of EYFS ages and stages.'* (Maths Champion survey)

Parental engagement

Several nurseries reported that parents of children in the intervention lacked mathematical knowledge themselves, meaning that they found it more difficult to support their children with their maths development. Thus, one of the perceived positive outcomes of the project was its resulting positive effect on parental confidence. The ability to engage and involve parents in their child's progress was thought to be an effective way to improve children's mathematics learning. Survey and case study responses showed that many nurseries made efforts to engage with parents, such as providing them with ideas for how to carry on MC style activities away from nursery.

> *'[Staff] speak with the parent at the end of that week to let them know how the children have got along […] and how they could extend it at home'.* (Maths Champion interview, Setting F, Best Practice)

This included 39% of nurseries sharing ideas for parents through social media or creating resource packs for parents to borrow and use at home to reinforce what had been done at nursery.

> *'We certainly fed back to parents quite a lot more about maths as a subject […] we will always share topic based ideas and how they can support learning at home'. 'I have just been really proud of the activities that [Champion] has set up for the children to take part in because they have inspired parents […] to not always believe that resources have to be particularly expensive'.* (Manager Interview, Setting E, Best Practice)

Staff felt that this increased effort led to more engagement and involvement from parents, who made an increased effort to use mathematical language and activities and to share this with nursery staff.

> *'Parents are more keen to share any home observations with key carers'* (Practitioner survey)

> *'I've noticed a lot of parents talking to their children about the time more when at nursery. Some even came in with watches.'* (Practitioner survey)

Although the two 'Best Practice' nurseries exemplified this communication with parents, not all nurseries were as successful at doing so. Parents were found to be too busy to talk to nursery staff, or their focus was often on literacy as opposed to maths.

> *'The parents come at half past seven and they drop their kids off and rush out the door'. 'The focus is usually on the concerns around speech, language, development and social concerns. From a parent's point of view, I don't think maths is pretty much on the radar'*. (Manager Interview, Setting B, End-point)

Other factors affecting perceived outcomes of the intervention

Research questions:	Dimensions:
5. What are the key success factors required for the Maths Champions intervention to work well?	• Quality • Monitoring of control/comparison conditions
6. What are the barriers to successful delivery of the intervention?	• Quality • Monitoring of control/comparison conditions

The process evaluation brought to light some evidence that testing introduced as part of the evaluation and not directly part of the Maths Champions intervention, may have had an indirect effect on the outcomes of the intervention in nurseries.

The ASPECTS assessment was used as the pre and post-intervention measure of attainment in the trial. Although opinions did differ, overall there was a large number of nursery staff who disliked the use of ASPECTS to assess the children. Staff expressed that it was not in-keeping with their pedagogy about child-led assessment or the MC style of holistic learning.

> '*A lot of [the children] seemed outside of their comfort zone in that conservatory that day with [an assessor] who was unfamiliar and using the laptop as the basis for a lot of investigative and exploratory maths activities'*. (Manager interview, Setting E, Best Practice)

However, this view was not held by every nursery. Settings C and D felt that ASPECTS was a useful tool to help measure progress and felt that it exposed strengths and weaknesses in the children that they had not noticed earlier.

> *'It was interesting to watch because some of the things we probably didn't realise they were capable of'*. (Manager interview, Setting C, Mid-point)

> *'The assessment stuff for the kids is very useful […] seeing which ones do struggle and stuff cos (sic) we thought some would be better and some would be worse'*. (Maths Champion interview, Setting D, Mid-point)

Formative Findings

There are two key barriers that need to be overcome in further iterations of the intervention if Maths Champions is to become more widely adopted. Firstly, the time and money available to nurseries in order to support the adoption and integration of the programme needs to be clearly identified and resourced accordingly. The process evaluation has provided some evidence that a lack of time and money had an impact on the ability of nurseries to successfully implement the intervention, and consequently there is a possibility that this may have impacted on its effectiveness. Secondly, the use and administration of the BKSB assessment should be considered. The process evaluation has provided some evidence that the BKSB test unintentionally resulted in a reported decrease in confidence in nursery staff. The project delivery team also reported in end of intervention period interviews that the relevance and importance of the BKSB component could now be reduced. If it is considered a truly necessary component of the intervention then it is important for implementers to understand its purpose, how it can be useful for the setting, and how to improve upon initial results. This may be facilitated by appropriate support from the project delivery team as part of the Maths Champions package.

Control Group Activity

As far as we are aware there was no resentful demoralisation in the control group and control group nurseries continued to deliver standard provision and did not drop out from the study as a result of their random allocation. However, no formal observations or interviews with control settings were conducted as part of the process evaluation.

Summary

Based on the views of nursery setting staff, the project delivery team and all additional information gathered through the process evaluation, there are some positive perceived outcomes as a result of the Maths Champions intervention, including reported increase in staff confidence, reported improvement in children's progress, and reported increase in parental engagement.

Conclusion

Key conclusions
1. Children who participated in Maths Champions made the equivalent of two additional months' progress in maths, on average, compared to children in the comparison group. This result has low security and should be treated with caution.
2. The impact of the intervention did not appear to be affected by pupils' eligibility for the Early Years Pupil Premium.
3. The trial found suggestive but inconclusive evidence that Maths Champions improved the quality of maths provision in participating nurseries.
4. A third of nurseries taking part in Maths Champions were very or partially engaged with all core features of the intervention.
5. Most nurseries were positive about Maths Champions and its impact on settings and children. However, some also raised the burden on nursery staff time as a potential issue.

Interpretation

The results of this large cluster randomised controlled trial find no statistically significant evidence that Maths Champions with additional support elements is effective in increasing the mathematical development and skills of children aged three and four years. The observed effect size is positive but small, equivalent to two months' additional progress, and not statistically significant (effect size 0.10, 95% CI -0.13 to 0.33, p=0.41). Similarly, no evidence was found that Maths Champions increases the language development and skills of children aged three and four years with a slightly larger but still not significant effect size of 0.17 (95% CI -0.06 to 0.40, p=0.15), equivalent to 2 months' additional progress. This was a pre-identified possible secondary impact of Maths Champions assessed via the ASPECTS.

The trial found suggestive but inconclusive evidence that Maths Champions with additional support elements improves the quality of maths provision in PVI nurseries and suggestive but inconclusive evidence that Maths Champions increases nursery practitioner confidence in their own maths ability and in helping nursery aged children learn maths, both of which were outcomes identified in the logic model as pre-cursers to improving children's mathematical development. Although there was a statistically significant, and potentially large effect on ECERS maths composite score (effect size 0.46, 95% CI 0.03 to 0.88), it is possible this finding is due to chance since effects on ECERS-3 and the maths subscale of the ECERS-E were non-statistically significant.

The trial was ultimately underpowered to detect a statistically significant effect size as small as 0.10, as was observed for the primary outcome. Whilst not statistically significant, the effect size based on ASPECTS mathematics score is positive and, since the Maths Champions intervention is inexpensive, a small but cost-effective benefit cannot be ruled out.

All results were robust to sensitivity analyses, including multiple imputation for the primary outcome of ASPECTS mathematics score. Despite the large amount of missing data (36% of pupils eligible and consenting at baseline were not included in the primary analysis model due to missing pre- and/or post-intervention ASPECTS data) results following multiple imputation were very similar to the complete case primary analysis model (effect size 0.09, 95% CI -0.06 to 0.23, p=0.38).

The effect of the intervention on the mathematical development and skills of children aged three and four years was not substantially altered by the average number of hours the child was reported to attend the nursery, having EYPP status, or gender.

Only a third of intervention nurseries were considered to have been very or partially engaged in the Maths Champions intervention (i.e. engaged in all identified core components). The CACE estimate of the effect of being very or partially engaged in *all* of the core aspects of the intervention on the pupils' mathematics attainment was a predicted increase of 1.09 points (95% CI -1.72 to 3.91, p=0.45). This equates to a moderate effect size of 0.22 (95% CI -0.34 to 0.78). Therefore, better compliance appears to result in a greater intervention effect; however, this increased effect is still not statistically significant.

This trial was designed and conducted to the CONSORT extension guidelines for cluster randomised controlled trials (Campbell et al, 2010). Importantly, independent concealed randomisation was used to ensure clusters were allocated without the possibility of bias. Post-testing was also conducted by independent blinded administrators, reducing the possibility of bias.

Based on the views of nursery setting staff and the project delivery team, and all additional information gathered through the process evaluation, there are significant positive perceived outcomes of the Maths Champions intervention, including reported increase in staff confidence, reported improvement in children's progress, and reported increase in parental engagement. During the evaluation period some additional support for nurseries allocation to Maths Champions was included in the form of regional induction events and face-to-face support visits which may have impacted on the number of nurseries who were very or partially engaged in *all* of the core aspects of the intervention.

Limitations

This trial suffered from high levels of nursery and child attrition: 353 (of 981, 36%) children from 14 of the 108 (13%) randomised nurseries (Intervention group n=10; Control group n=4) could not be included in the primary analysis. The reasons provided by nurseries for withdrawing from the trial do not suggest a specific link with the intervention. A large number of children were lost to follow up due to leaving the nursery between baseline and post-testing (n=187; Intervention group n=108; Control group n=79), and being absent on the day of post-testing (n=44; Intervention group n=28; Control group n=16). Although some attrition on this basis was anticipated, the amount was greater than expected and movement between, and removal from, institutions of children in this age group should be considered in future sample size calculations for trials involving children attending nursery settings. However, we do not think that these post-randomisation exclusions are likely to have introduced bias as there is no reason to suppose that their loss was linked to the intervention. Reassurance for lack of selection bias due to attrition is given in the baseline table (Table 7) where there is little difference between the analysed groups: this suggests that the effect of attrition on observed variables was similar between the two groups, implying that selection bias was not a factor. Additionally, our complete case and MI results are very similar, adding weight to this premise.

The original sample size was based on randomising 120 nurseries to have 80% power to detect an effect size of 0.20 of a standard deviation between the control and intervention groups, allowing for 10% attrition at the child level, and assuming an intracluster correlation coefficient (ICC) of 0.19[5], 10 children per nursery and a pre and post-test correlation of 0.70 (CEM, 2001). Based on the number of pupils included in the primary analysis model, and the observed ICC (0.17) and the correlation between the pre- and post-test ASPECTS mathematics scores (0.59), the estimated minimum detectable effect size was 0.26, which is larger than the observed effect size. The trial was therefore underpowered to detect a statistically significant effect size as small as 0.10.

[5] Based on ICC observed in Every Child Counts (ECC) evaluation. Torgerson C.J., Wiggins A., Torgerson D.T., Ainsworth H., Barmby P., Hewitt C., Jones K., Hendry V., Askew M., Bland M., Coe R., Higgins S., Hodgen J., Hulme C. & Tymms P. (2011). The Every Child Counts Independent Evaluation Report. Department for Education.

In regards to generalisability all nurseries included in this trial were PVI nurseries and every effort was made to recruit nurseries serving disadvantaged children, with 65 (60.2%) randomised nurseries being located in the 50% most deprived super output areas (as determined by the IDACI 2015). Some children included in the evaluation did not meet the pre-specified inclusion criteria, 74 children tested at baseline were aged less than 36 months (Intervention group 34/407 (8.4%); Control group 40/438 (9.1%)); however, all were aged over 30 months and turned three during the evaluation. It is possible that some children included in the evaluation attended nursery for less than the required 15 hours a week. Information collected at post-test suggested 46 children attended for less than 15 hours a week (Intervention group 19/407 (4.7%); Control group 27/438 (6.2%)).

Future research and publications

This is a low cost intervention, although this effectiveness trial found no statistically significant evidence of positive effects, potentially small beneficial effects cannot be ruled out. Any future research investigating the effectiveness of Maths Champions on increasing the mathematical development and skills of children aged 3 and 4 would need to be powered to detect a small effect size.

This research did not explore implementation of Maths Champions beyond the first year. It is suggested that follow up of nurseries participating in the evaluation in the two years following the initial one year programme would be useful to understand whether settings continue with the Maths Champions approach and to learn about the sustainability of the programme.

References

Altman, D.G., & Bland, J.M., (2005). Treatment allocation by minimisation. *BMJ*, 330 :843.

APPG Maths and Numeracy (2014). Maths and numeracy in the early years. *www.nationalnumeracy.org.uk/sites/default/files/.../appg_paper_-_eys.pdf*

Campbell M.K., Piaggio G., Elbourne D.R., Altman D.G.; for the CONSORT Group., (2010). Consort 2010 statement: extension to cluster randomised trials. *BMJ*. 345:e5661. PMID: 22951546.

Curriculum Evaluation and Management Centre (CEM) (2001) *Assessment Profile On Entry For Children And Toddlers* , Pub. CEM Centre, University of Durham (http://www.leeds.ac.uk/educol/documents/00002299.htm).

Chen, Jie-Qi., McCray, Jennifer., Adams, Margaret., Leow, Christine. (2014) A Survey Study of Early Childhood Teachers' Beliefs and Confidence about Teaching Early Math *Early Childhood Education Journal* (2014) 42:367–377 DOI 10.1007/s10643-013-0619-0

Department for Education (2015). Source: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/483897/SFR47_2015_text.pdf/ Last Accessed 8th January 2018

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. *Stat Methods Med Res*, 14(4): p. 369-95.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P.,& Japel,C. (2007). School readiness and later achievement. *Developmental psychology*, *43*(6), 1428 -1446.

Easton, C., & Gee, G. (2012*). Early intervention: informing local practice* (LGA Research Report). Slough: National Foundation for Educational Research. https://www.nfer.ac.uk/publications/LGLC02/LGLC02.pdf

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide* (NCEE 2014-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: http://whatworks.ed.gov

Hedges, L.V. (1981).Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics,* 6: 107-128

Hedges, L. (2007). Effect Sizes in Cluster-Randomized Designs. *Journal of Educational and Behavioral Statistics*, Vol. 32, No. 4, pp. 341-370

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. & Kerr, K. (2016). Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook. Available at: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Guidance_Final.pdf

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Developmental psychology*, *45*(3), 850.

Mathers, S & Smees, R. (2014). *Quality and Inequality: Do three- and four- year olds in deprived areas experience lower quality early years provision?* A Nuffield Foundation Report.

Mathers, S., Ranns, H., Karemaker, A., Moody, A., Sylca, K., Graham, J. & Siraj-Blatchford, I. (2011). *Evaluation of the Graduate Leader Fund. Final Report. Research Report DFE-RR144.* Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/181480/DFE-RR144.pdf

Melhulish, E., & Gardiner, J. (2017). Study of Early Education and Development (SEED): Study of Quality of Early Years Provision in England. Research Report, DfE; London.

NDNA (2014) *NDNA project report: Maths Champions 2013-2014* Copyright NDNA

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. Review of Educational Research, 78, 33–84.

Saghaei, M. and Saghaei, S. (2011) Implementation of an open-source customizable minimization program for allocation of patients to parallel groups in clinical trials. *Journal of Biomedical Science and Engineering*, 4, 734-739. doi: 10.4236/jbise.2011.411090. Available at: http://www.scirp.org/journal/PaperInformation.aspx?PaperID=8518

Sammons, P., Sylva, K., Melhuish, E.C., Siraj-Blatchford, I., Taggart, B., Smees, R., Draghici, D. and Toth, K. (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.

Schulz, K.F., Altman D.G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*, 152(11): p. 726-32.

Senn, S. (1994). Testing for baseline balance in clinical trials. *Stat Med*, 13(17): p. 1715-26.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *The Effective Provision of Pre School Education (EPPE) Project: Final Report A Longitudinal Study Funded by the DfES 1997-2004*. London, UK.

Sylva, K., Melhuish, E., Sammons, P., Siraj, I,. Taggert, 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*: September 2014.

Williams, P. (2008). *Independent review of mathematics teaching in early years settings and primary schools.* A report for the Department for Education.

Appendix A: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found here. Cost ratings are awarded as follows:

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

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	Power	Attrition[6]					
5 🔒	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%			Adjustment for Balance [0]		
4 🔒	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%					
3 🔒	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%					
2 🔒	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%	2		Adjustment for threats to internal validity [0]		2
1 🔒	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	41-50%					
0 🔒	No comparator	MDES > 0.6	over 50%					

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = 2 padlocks as pupil level attrition was 36%, power at randomisation was 0.22 and it was a well-conducted experimental design.
- **Reason for adjustment for balance** (if made): Pre-test differences for the main outcome were equivalent to an effect sizes of 0.09. Thus, an adjustment of 1 padlock could be made. As this trial presents no further threats to validity and the robustness checks accounting for missing data showed very similar results, no further adjustments for imbalance are recommended.
- **Reason for adjustment for threats to validity** (if made): None found.
- **Final padlock score:** initial score adjusted for balance and internal validity = 2 padlocks

[6] Attrition should be measured at the pupil level (even for clustered trials) and from the point of randomisation to the point of analysis.

Additional Appendices:

Additional appendices are available in a separate document, which includes:

- **TIDieR checklist**
- **The Maths Champions Intervention: Theory of change**
- **Pilot – Maths Champion Survey**
- **Pilot – Practitioner Survey**
- **Pilot – Nursery Information Sheet**
- **Pilot – Nursery Consent Form/MOU**
- **Pilot – Parent Information Sheet**
- **Pilot – Parent Consent Form**
- **Effectiveness Trial – Nursery Information Sheet and Memorandum of Understanding**
- **Effectiveness Trial – Information Sheet for Parents**
- **Effectiveness Trial – Parent FAQs**
- **Effectiveness Trial – Consent Form for Parents**
- **Effectiveness Trial – Parent Consent for Researcher Assessment at Baseline**
- **Effectiveness Trial – Reminder Letter to Parents**
- **Effectiveness Trial – Maths Champion/Practitioner Confidence Survey**
- **Effectiveness Trial - Maths Champion Post-Intervention Survey**
- **Effectiveness Trial - Practitioner Post-Intervention Survey**
- **Effectiveness Trial Interview Information Sheet and Consent Form**
- **Effectiveness Trial Start-point interview schedules**
- **Effectiveness Trial Mid-point interview schedules**
- **Effectiveness Trial End-point interview schedules**
- **Effectiveness Trial Best Practice interview schedules**
- **Effectiveness Trial Maths Champion post intervention survey findings**
- **Effectiveness Trial Practitioner post intervention survey findings**
- **Practitioner Confidence Survey Results**

This document is available for download at **www.educationendowmentfoundation.org.uk**

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk