Statistical Analysis Plan
for Maths Champions
Evaluating Institution: York Trials Unit

Education Endowment Foundation

.INTERVENTION	Maths Champions Programme
DEVELOPER	National Day Nurseries Association & Oxford University
EVALUATOR	Durham University, The York Trials Unit (YTU)
TRIAL REGISTRATION NUMBER	13051035
TRIAL STATISTICIAN	Caroline Fairhurst
TRIAL CHIEF INVESTIGATOR	Hannah Ainsworth and Victoria Menzies (maternity cover Lyn Robinson-Smith)
SAP AUTHOR	Caroline Fairhurst
SAP VERSION	1.0
SAP VERSION DATE	07/12/2017

This analysis plan was written post-randomisation and prior to receipt of any outcome data, and deals only with the statistical analysis of effectiveness for the main trial. This document has been written based on information contained in the study protocol version 3 dated 17th January 2017 in which full details of the background and design of the trial are presented.

Protocol and SAP changes

Any changes made to the protocol which impact on the SAP, and any changes made to the SAP after its initial publication, will be specified here. There are no such changes to note to date.

Changes from current version of the protocol included in this SAP

- The language component of the ASPECTS assessment will be 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 in this document.

- Details of a post-hoc power calculation have been included based on the number of nurseries and children recruited into the trial.

- 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.

- In the protocol, we state that within analysis models adjustment will be made for clustering within nurseries using robust standard errors; however, since the EEF guidance on analysis recommends the use of hierarchical (mixed) models, we have amended the proposed analysis model to a mixed model with adjustment for clustering using nursery as a random effect.

- In addition to baseline assessment, regression models for the ASPECTS will also be adjusted for the factors used in the minimisation (EMA, 2015).

- In the protocol, we proposed to consider adjusting the ECERS analyses by mean childcare qualification level of staff at the nursery. In the event, this level of details will not be available,

and so we shall adjust instead for the proportion of paid childcare/teaching staff at the nursery that are qualified to Level 3 or above.

- On the advice of Sandra Mathers from the University of Oxford, who is assisting with the collection of ECERS data, we have added in an additional maths-related secondary outcome related to the ECERS. This is a composite maths score that combines the mathematics-related items of the ECERS-3 and the ECERS-E maths subscale.

- A secondary intention-to-treat analysis is detailed for the primary outcome. In line with EEF policy, this secondary analysis will repeat the primary ITT model, but will adjust only for group allocation and prior attainment as fixed effects (and nursery as a random effect), i.e. minimisation factors will be omitted from this model. This is to ease comparability across all EEF trials for work Durham University is undertaking on the overarching evaluation of the EEF.

Table of Contents

Introduction

Maths Champions is a one-year programme developed by the National Day Nursery 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. This two armed cluster randomised controlled trial with randomisation at the nursery level will evaluate the effectiveness of the Maths Champions Programme on the mathematical development and skills of children aged three and four years. Although all children within the nursery will receive the intervention, the primary outcome of the evaluation will focus on the mathematical attainment of children who are aged three at the start of the intervention, due to attend school in September 2017 and attend nursery for a minimum for 15 hours per week. Secondary research questions include:

1. How effective is the Maths Champions intervention at improving nursery practitioners' confidence in supporting children's maths development? [Secondary outcome 1]

2. 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]

3. 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]

Study design

This is a pragmatic two armed cluster randomised trial with randomisation at the nursery level. Nurseries were randomly allocated 1:1 to either receive the Maths Champions intervention (a one-year intervention); or to continue with usual nursery provision (control). There are two assessment time points in this trial – baseline (June/July 2016), and post-intervention (June/July 2017). Participating children at the nurseries will be assessed using the Assessment Profile on Entry for Children and Toddlers (ASPECTS) at baseline and post-intervention a year later, just before they leave the nursery. Nursery-level assessment using the Early Childhood Environmental Rating Scales 3 (ECERS-3) and the Early Childhood Environmental Rating scale extension (ECERS-E) will take place post-intervention, and nursery staff will complete a practitioner confidence survey post-intervention. Further details are given in the section below on Outcome Measures.

All nurseries (both intervention and control) will be provided with a financial contribution of £500 to acknowledge the time and resources required to take part in the trial (provided by NDNA).

Private, Voluntary and Independent (PVI) nurseries whose child population includes children who are three years old, located in areas of high deprivation (participating nurseries located in the top 40% most deprived areas as determined by the Income Deprivation Affecting Children Index (IDACI)) and who haven't previously been involved with the Maths Champions intervention were eligible to participate in the trial. At the start of the trial, it was estimated that there were approximately 60 nurseries currently implementing Maths Champions in the UK. NDNA ceased registration for the Maths Champions intervention in December 2015, to ensure capacity was sufficient to support the project. Recruitment of nurseries began in February 2016. Nurseries were only eligible to take part in the study if they agreed to all of the study requirements outlined in the Memorandum of Understanding with Nurseries.

Children aged three years old at the start of the intervention, due to start school in September 2017, and attending nursery for a minimum of 15 hours per week were eligible to participate in the trial for the purposes of evaluation (all children in a participating setting may have been exposed to the intervention). The inclusion criteria related to age and minimum number of hours was to ensure that children in the evaluation are in the nursery long enough to receive three academic terms of the intervention. Parents/carers of eligible children were informed of the evaluation via a letter passed on to them by the nursery and were required to give opt-in consent: i) for their child to participate in the

baseline and outcome testing; ii) to be contacted should school destination data for their child not be available from the nursery; and iii) for long term tracking of their child's educational attainment via the NPD. Parents/carers could choose to opt-in to the assessment, but not to the follow-up or for their child's data to be linked to the NPD.

Randomisation

In July 2016, the independent trial statistician, who was not involved in the recruitment of nurseries, used a dedicated computer program, MinimPy, 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 2015/16 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. 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).

Calculation of sample size

From protocol
We make the following assumptions: an intra cluster correlation 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 (i.e., 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 of Education.

At randomisation

The final number of nurseries randomised into the trial was 108 (Intervention 54; Control 54); however, 8 dropped out between randomisation and completion of the baseline testing. In total, 888 children were assessed using ASPECTS at baseline across 100 nurseries; however, to date 2 nurseries have yet to return any parental consent forms to use their child's data to the trial team. Therefore, 98 nurseries have currently returned all or some of the consent forms to the evaluation team, for 854 children. We are actively encouraging the return of the outstanding 34 parental consent forms from nurseries. Therefore, 854 children from 98 nurseries for whom we have parental opt-in consent were assessed for baseline ASPECTS. With this number, assuming an average of 9 (854/98=8.7, rounded to 9) children per nursery, an ICC of 0.19, a pre-posttest correlation of 0.70 and 10% loss to follow-up, we will have 80% power to detect an effect size of 0.23 between the two arms.

Outcome measures

The primary outcome is maths attainment at the end of nursery measured using the ASPECTS, which was produced by the Centre for Evaluation and Monitoring (CEM) at Durham University. Participating children were assessed using ASPECTS at baseline and will be 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. The 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.

Within each nursery, testing was conducted on approximately ten children who met the eligibility criteria and whose parents provided the relevant consent, in order to keep the testing burden on the nurseries to a minimum. At baseline, the children were tested by their nursery practitioner. At follow-up, they were assessed by an independent, blinded research assistant trained by the evaluation team. If there were less than ten children per setting, all eligible children were tested. If there were more than ten eligible children then up to 14 children were randomly selected for testing. This was done by members of the trial team by generating a random number for each child, ordering the children on these numbers and selecting from the top down the number of children required.

Other secondary outcomes include practitioner confidence, gathered via a short survey, adapted from the version used by Chen at al (Chen, 2013). We requested this would be completed by a graduate within the setting and all of the practitioners who worked with children aged 3+. The survey is completed on paper at post-intervention only and consists of three subscales: Beliefs About Nursery Aged Children and Maths (8 items); Confidence in Helping Nursery Aged Children Learn Maths (11 items); and Confidence in Own Maths Abilities (9 items). Each item is 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 can be summed to produce three summary scores (Beliefs About Nursery Aged Children and Maths: scored from 8 to 40; Confidence in Helping Nursery Aged Children Learn Maths: 11 to 55; Confidence in Own Maths Abilities: scored 9 to 45). The three sections represent three different constructs and so it is not possible to combine them into a total score.

The Early Childhood Environmental Rating Scales 3 (ECERS-3) and the Early Childhood Environmental Rating scale extension (ECERS-E) will be collected at the nursery level at the end of the trial in both intervention and control nurseries, in July/July 2017 following the intervention period (or an equivalent time for the control group). Both of these scales 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. A summary score for the ECERS-3 is computed by summing the item scores and dividing by 35 to obtain a total mean score from 1 to 7. The collection of ECERS-E data will focus on the following maths subscale including the 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 score for the ECERS-E maths 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 from 1 to 7. The collection of these data requires a full day's observation within each nursery setting which will be conducted by and subcontracted to company A+ Education Ltd who will subsequently provide the data to the evaluation team for analysis. Assessors will be blinded to random allocation. Observers will complete the ECERS-3 in a 3-hour morning observation, and the ECERS-E across the whole day (a 6-hour observation). Additionally, so that no information on maths practice is lost, observers will also complete the three maths-related items of the ECERS-3 again in the afternoon. This information will be used to create an additional secondary outcome of a composite maths score, calculated as follows: i) take the mean of the maths ECERS-3 items assessed in the morning; ii) take the mean of the maths ECERS-3 items assessed in the afternoon; iii) take the mean of these two means; and iv) add this mean to the ECERS-E maths subscale score.

Other important information

Nurseries were asked to provide data, for each child, on their date of birth, gender, early years pupil premium (EYPP), ethnicity and average number of hours the child attends the nursery per week.

Compliance and fidelity

Each nursery in the intervention arm will be assessed for their implementation fidelity (the extent to which they implemented Maths Champions as intended by NDNA). The NDNA will rate each setting on eightcompulsory 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). The ratings on these eight components will be summed out of a possible 0-16. The definitions are outlined in the table below.

Compulsory Components Fidelity Rating

Criteria	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

Criteria	Description	RAG rating
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

Analysis

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

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

The trial has been designed and conducted, and will be reported to, CONSORT standards (Schulz et al, 2010; Campbell et al, 2012) in order to minimise all potential threats to internal validity, such as selection bias and a range of post randomisation biases (Cook and Campbell, 1979; Shadish, Cook and Campbell, 2002; Torgerson and Torgerson, 2008). In this way, unbiased estimates of the impact of the intervention will be provided. A CONSORT diagram will be produced 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, the number selected to take part in the evaluation, and the numbers actually tested for ASPECTS at baseline and follow up will be reported with reasons for non-participation given where available.

Nursery withdrawals will be classified as: "withdrawn pre-randomisation", "nursery no longer wishes to receive intervention but agrees to outcome data being collected", and "complete withdrawal from intervention and data collection". The numbers (with reasons and timings) of losses to follow-up (drop-outs and withdrawals) over the course of the trial will be summarised by treatment arm.

Primary intention-to-treat (ITT) analysis

Numeracy attainment for children in the intervention group and those in the control group will be compared using a linear mixed model at the child-level. Group allocation, baseline ASPECT numeracy score, and nursery-level minimisation factors (number of graduate staff and number of children leaving for primary school in September 2017) will be included as fixed effects in the model. The continuous variables that were dichotomised to use as factors in the minimisation will be included in their continuous form in the model.

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

Secondary ITT analysis

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

Interim analyses

No interim analyses will be undertaken.

Imbalance at baseline for analysed groups

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

Missing data

The amount of missing baseline and primary outcome data will be summarised, and reasons for missing data explored and provided in the report, where available. The impact of missing data on the primary analysis will additionally be assessed using multiple imputation by chained equations, where greater than 5% of children with baseline ASPECTS data are missing from the primary analysis model due to missing outcome and/or covariate data. Pre- and post-intervention ASPECTS mathematics score data will be predicted by a linear regression model that includes all variables included in the primary analysis model.

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

Non-compliance with intervention

A Complier Average Causal Effect (CACE) analysis (Dunn, 2005) for the primary analysis (as described in the "Primary intention-to-treat (ITT) analysis" section) will also be considered to account for compliance/engagement of the nurseries with the intervention. The NDNA will be able to provide 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 will be summarised. Formal CACE analyses for the primary outcome at the child-level to investigate the effects of compliance will be 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 compulsory 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 and optional components of the intervention are 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 compulsory aspects of the intervention i.e. all core components 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

The language score from the ASPECTS will be analysed in the same way as the primary outcome as it is described in the section: Primary intention-to-treat (ITT) analysis.

Scores for ECERS-3, the maths subscale of the ECERS-E and the ECERS maths composite score will be compared between the nurseries in the two groups using linear regression. There will not be a baseline value for the nurseries to include in the analysis; the models will adjust 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).

Responses to items in the practitioner confidence survey will be summarised descriptively by trial arm. These will be presented for just the 'Maths Champion' of the nursery (where this person can be identified), and for all respondents The three subscale scores will be compared between the two 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. These will be undertaken just for the responses for the Maths Champions (one per nursery), and again, separately, using all respondents by including nursery as a random effect.

Additional analyses

During the course of the trial, two nurseries (one intervention and one control) amalgamated. The children in these two nurseries will be 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 propose that one assessment of the environment is made and the resulting score for the ECERS assigned to both nurseries for the main analysis. In a sensitivity analysis, we will repeat 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.

The intra-cluster correlation coefficient (ICC) associated with nursery for the primary outcome (both pre and post-test) will be presented alongside a 95% CI. The correlation between baseline and follow-up ASPECTS scores will be calculated using Pearson's correlation coefficient.

Subgroup analyses

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

Effect size calculation

Effect sizes will be calculated based on the adjusted mean difference between the intervention and control group (controlling for prior attainment and the minimisation factors) and the total variance (between plus within nursery variance), obtained from the multi-level model. These two figures will be reported. We will also report the effect sizes for the secondary analyses and the sub-group analyses following the same principles. The effect size and 95% confidence interval will be calculated using equations (19) and (20) given in Hedges (2007) for cluster randomised designed analysed via multilevel models and allowing for unequal cluster sizes.

Report tables

The EEF trial report template contains several tables whose structure is pre-specified. These tables have been adapted as appropriate for this trial and included below, along with other results tables that will be included in the report.

Example Table: Summary of impact on primary outcome

Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	EEF cost rating
Intervention vs. control				
Sensitivity analysis with multiple imputation: Intervention vs. control				
Sensitivity CACE analysis: Intervention vs. control				

Example Table: Minimum detectable effect size for primary outcome

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	120 (60; 60) 1,200 (600; 600)	0.7[a]	0.19[a]	80%	0.05	0.20
Randomisation		[a]	[a]			
Primary analysis		[b]	[b]			
[a] based on assumptions; [b] actual observed correlation and ICC						

Example Table: Baseline comparison for participating nurseries

Variable	Intervention group		Control group	
Nursery-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
N of graduate staff 1 >1				
N of children leaving for primary school in Sept 2017 <22 ≥22				
Nursery-level (continuous)	**n (missing)**	**Mean (SD)** **Median (min, max)**	**n (missing)**	**Mean (SD)** **Median (min, max)**
N of graduate staff				
N of children leaving for primary school in Sept 2017				
Childcare qualification level of staff				
% of non-white British children				

Example Table: Baseline comparison for children at randomisation

Variable	Intervention group		Control group	
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for Early Years Pupil Premium				
Sex, male				
Ethincity				
Pupil-level (continuous)	**n (missing)**	**Mean (SD)** **Median (min, max)**	**n (missing)**	**Mean (SD)** **Median (min, max)**
Baseline ASPECTS Mathematics Language				
Hours attendance at nursery per week				

Example Table: Baseline comparison for children included in the primary analysis

Variable	Intervention group		Control group	
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
Eligible for Early Years Pupil Premium				
Sex, male				
Ethincity				
Pupil-level (continuous)	n (missing)	Mean (SD) Median (min, max)	n (missing)	Mean (SD) Median (min, max)
Baseline ASPECTS Mathematics Language				
Hours attendance at nursery per week				

Example Table: Summary statistics and results from the analysis models for pupil-level outcomes

	Raw means/n (%)				Effect size		
	Intervention group		Control group				
Primary outcome	n (missing)	Mean (SD) Median (min, max)	n (missing)	Mean (SD) Median (min, max)	n in model (intervention; control)	Effect size (95% CI)	p-value
ASPECTS mathematics							
Sensitivity analysis (multiple imputation)	n (missing)	Mean (SD) Median (min, max)	n (missing)	Mean (SD) Median (min, max)	n in model (intervention; control)	Effect size (95% CI)	p-value
ASPECTS mathematics							
Secondary outcome (continuous)	n (missing)	Mean (SD) Median (min, max)	n (missing)	Mean (SD) Median (min, max)	n in model (intervention; control)	Effect size (95% CI)	p-value
ASPECTS language							

Example Table: Summary statistics and results from the analysis models for nursery-level outcomes

	Raw means/n (%)				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (SD) Median (min, max)	n (missing)	Mean (SD) Median (min, max)	n in model (intervention; control)	Effect size (95% CI)	p-value
ECERS 3							
ECERS E mathematics							
Sensitivity analysis: ECERS 3							

Sensitivity analysis: ECERS E mathematics						

Example Table: Summary statistics and results from the analysis models for practitioner-level outcomes

Outcome	Raw means/n (%)				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (SD) Median (min, max)	n (missing)	Mean (SD) Median (min, max)	n in model (intervention; control)	Effect size (95% CI)	p-value
Beliefs About Nursery Aged Children and Maths							
Confidence in Helping Nursery Aged Children Learn Maths							
Confidence in Own Maths Abilities							

References

Altman DG, Bland JM. Treatment allocation by minimisation. BMJ. 2005;330(7495):843.

Chen, J.-Q., et al., A Survey Study of Early Childhood Teachers' Beliefs and Confidence about Teaching Early Math. Early Childhood Education Journal, 2014. 42(6): p. 367-377.

Campbell MK, Piaggio G, Elbourne DR, Altman DG; for the CONSORT Group. *Consort 2010 statement: extension to cluster randomised trials.* BMJ. 2012 Sep 4;345:e5661. PMID: 22951546

Cook, T.D. & Campbell, D. (1979). *Quasi-experimentation: Design and Analysis Issues for Field Settings.* Boston: Houghton Mifflin.

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).

Dunn, G., M. Maracy, and B. Tomenson, *Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods.* Stat Methods Med Res, 2005. 14(4): p. 369-95.

European Medicines Agency, *Adjustment for baseline covariates in clinical trials.* 2015.

Hedges, L. (2007). Effect Sizes in Cluster-Randomized Designs. *Journal of Educational and Behavioral Statistics, 32*(4), 341-370. Retrieved from http://www.jstor.org/stable/20172092

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.* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/181480/DFE-RR144.pdf

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

Shadish, W.R, Cook, T.D & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA, US: Houghton, Mifflin and Company.

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

StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC

Torgerson, D.J. & Torgerson, C.J. (*2008). Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction. Palgrave Macmillan.*