Family Skills Analysis Plan

NatCen Social Research

INTERVENTION	Family Skills
DEVELOPER	Learning Unlimited, Campaign for Learning
EVALUATOR	NatCen Social Research
TRIAL REGISTRATION NUMBER	ISRCTN90043546
TRIAL ANALYSTS	Karl Ashworth, Robert Wishart
TRIAL CHIEF INVESTIGATOR	Karl Ashworth / Fatima Husain
SAP AUTHOR	Martina Vojtkova, Nico Jabin
SAP VERSION	4.0
SAP VERSION DATE	04 September 2017
EEF DATE OF APPROVAL	13.09.17
DEVELOPER DATE OF APPROVAL	23.08.17

Table of Contents

1. Introduction

This document sets out the detail of the analysis planned for the cluster-randomised controlled efficacy trial of the Family Skills intervention.

Family Skills (Supporting Kids in Literacy, Learning and School) is a family literacy intervention for parents/carers and children for whom English is an additional language (EAL). The programme is implemented by 16 delivery partners[1] in England and comprises parents and their Reception-aged children receiving 30 hours of family literacy support in school, with parents expected to conduct follow-up activities at home.

The intervention consists of 11 two-and-a-half hour sessions, delivered over one school term (three months). The first two hours of each session involve parents only; the final half hour involves parents and children learning together[2]. With a focus on how to make most of bilingualism, the sessions cover a range of topics including:

- an introduction to education in England
- the culture of schools
- reading strategies and phonics
- home literacy practices
- oral traditions (including storytelling, songs and rhymes)
- strategies for learning through play.

The sessions are delivered by qualified Adult/Family Learning tutors, employed by the local delivery partner. The tutors receive one full day of training about the Family Skills intervention, and a Family Skills Toolkit. The Family Skills Toolkit provides session plans, resources and activities for the intervention, as well as resources for recruiting parents to take part in the programme. The activities and resources are sourced from pre-existing best practice and compiled in consultation with practitioners involved in Family Learning and EAL. Additional intervention activities include a school tour, a visit to a local library, and a talk about phonics.

Delivery partners are responsible for recruiting schools into the trial, and for recruiting parents to participate in the Family Skills programme. In practice, family recruitment practices differ across sites and are led either by management-level staff from the delivery partner, by the tutor, and/or by the school. The delivery partners coordinate with the school about session timetabling and deliver the sessions.

The Family Skills intervention has been designed to support families in developing their children's English and literacy skills. It does this in three ways, by

- equipping parents with greater knowledge of how their children are taught to read
- developing parents' English language skills, and
- acquainting parents with strategies and activities to support their children's literacy development at home.

These new skills and knowledge are expected to increase parents' confidence and allow closer engagement in their children's learning. Ultimately, this should lead to improvements in literacy among children.

---

[1] The delivery partners are made up of a range of organisations providing adult learning services, including local authorities. The full list is given in Table 2 on page 9.
[2] With the exception of one session which is entirely parents-only.

Primary hypothesis

The primary hypothesis (**H1**) is that eligible reception-year pupils whose parents/carers have been assigned to receive an offer to participate in the *Family Skills* intervention will on average have, six months after the intervention's start and around three months after its end, better literacy outcomes than eligible pupils that have been assigned to the control group and who have been offered business-as-usual support for their literacy skills development.

Secondary hypotheses

Five secondary hypotheses will be examined, as set out below. These hypotheses will be considered preliminary and indicative.

On-treatment analysis:

- **H2**: Pupils whose families participate in *Family Skills* will have a better literacy outcome than pupils of non-participating eligible families at three-month follow-up.

Sub-group effects:

- **H3**: *Family Skills* will have a different (higher or lower) impact on pupils participating in the intervention and eligible for the pupil premium than those who did not participate in the intervention.
- **H4**: *Family Skills* will have a different (higher or lower) impact on female and male pupils participating in the intervention.
- **H5**: *Family Skills* will have a different (higher or lower) impact on pupils with different baseline English language skills (assessed using CEM's BASE[3] test).

Additional analyses:

- **H6:** The number of Family Skills sessions attended by the parent/carer will be positively correlated with participating pupils' literacy outcomes, controlling for other factors.

2. Study design

Population, eligibility criteria and sample

Pupils eligible to participate in the programme are all reception-year pupils identified by schools as having English as an additional language (EAL). We are aware that there might be a diversity of definitions used by schools to identify EAL pupils. However, since the programme is aimed at EAL pupils as defined by schools, we believe that this is a suitable way to identify the target population.

Pupils participating in the trial include all eligible pupils that have not opted out of the research and whose literacy skills have been assessed by the participating schools at baseline. Due to challenges encountered during baseline testing, not all eligible pupils have been assessed at baseline. Where these challenges were identified during the baseline testing period and the school had more than one form entry in reception year, we have randomly sampled forms to be assessed[4]. This approach was taken to

[3] http://www.cem.org/reception-baseline-assessment

[4] In these instances, the NatCen evaluation team determined a random ordering of forms and agreed with the relevant school that they will test as many forms as possible within the available time and resources according to the order provided by the NatCen team.

minimise the potential risk of bias that could be introduced by schools selecting which pupils to test and therefore include in the trial.

All families with an eligible child are eligible to participate in the programme, regardless of whether the family consented for their child to participate in the research or whether attainment was assessed at baseline. However, only families that have consented for their child to participate in the research will be included in the evaluation[5].

Delivery partners were asked to recruit schools on the basis that they have higher than average proportions of pupils with EAL[6] and a minimum of two form entry, in order to maximise the numbers of families opting into the trial. However, due to challenges in school recruitment, eligibility criteria were revised in agreement with the developers and EEF and all schools with a minimum of 6 pupils with EAL (including single form entry schools) were eligible to take part in the programme and trial.

Each delivery partner aimed to recruit 10 schools in their local area in England, covering 11 regional authority areas: Sussex, Buckinghamshire, South Yorkshire, Lancashire, Leicestershire, Northamptonshire, Nottinghamshire, Portsmouth, Hampshire, Birmingham and Greater London. The final number of successfully recruited schools that have completed baseline testing varied considerably between delivery partners, ranging from 2 to 12 schools[7]. The final sample consists of 115 primary schools recruited by 16 delivery partners.

Description of trial design

The evaluation is run as a cluster randomised controlled trial. Schools have been chosen as the unit of treatment and randomisation because the school is the main unit of intervention delivery. Moreover, school-level randomisation avoids potential spillover effects which can bias effect estimates.

Our aim is to test the effectiveness of the intervention in its ‘real world’ context by which we mean that schools’ choices and cultural and environmental factors are likely to have an impact on the extent to which an intervention can achieve its optimal outcome.

Description of trial arms

The schools participating in the trial were randomly assigned (by NatCen) into one of two intervention conditions:

- **Group 1** (intervention schools) will receive the Family Skills programme

[5] 78% of eligible pupils in treatment schools have been assessed at baseline. As a result, there are up to 22% of families whose children have not been assessed at baseline and who have not given consent who are nonetheless receiving the programme. These are not considered part of the study sample.
[6] With the average proportion of EAL pupils defined as 18% based on 2013 National Statistics (accessed at: http://www.naldic.org.uk/research-and-information/eal-statistics/eal-pupils/)
[7] The original recruitment target, based on an assessment of required trial power at protocol stage, was 140 schools to be recruited by 14 delivery partners. However, due to challenges and delays in school recruitment and concerns over low programme take-up limiting trial power, the EEF, developers and NatCen have agreed to the recruitment of two additional delivery partners and a total recruitment target of 155 schools. This target was successfully reached following an extension of the school recruitment period into the beginning of the autumn term 2016. Forty schools subsequently dropped out of the trial prior to, or over the course of baseline data collection, primarily driven by a misalignment of project and evaluation timelines. Reasons for school drop-out included lack of support for participation in the programme or trial among school leadership, schools not being fully aware about the requirements of the project when agreeing to participate short timelines for the completion of baseline data collection, and lack of capacity or other challenges in completing the baseline assessment.

- **Group 2** (control schools) will continue with 'business as usual'

Business as usual may include interventions and support to help Reception year pupils with EAL. The evaluation will not stipulate that control schools cannot provide certain support to these pupils. The business as usual control allows the impact evaluation to specify the effect of Family Skills on the literacy learning of pupils EAL relative to what schools are currently doing.

The use of a 'business as usual' control means that pupils in control schools are not made worse off as a result of participating in the trial compared to those in intervention schools, as they will not be denied access to existing support offered by schools (other than access to the Family Skills programme). There are no ethical reasons why control schools should be offered the intervention as part of a wait list approach until we know that the programme is effective and cost-effective in achieving the intended outcomes. Instead, control schools will be offered a financial incentive for taking part in the evaluation. This incentive will help to manage the risk of 'resentful demoralisation' and 'compensatory rivalry' of schools.

There is a small risk that longer-term (2-year follow-up, post-trial) differences between treatment and control groups do not reflect the true effect of Family Skills, since all participating schools will be given the option to implement Family Skills in the school year following the trial. Although implementation will focus on a different year group, children in the control cohort could receive intervention benefits if their parents take up the intervention for younger siblings. Overall, we consider this risk to be low because:

- not all control and treatment schools will choose to take up the programme in subsequent years
- the number of children in the treatment and control cohort with siblings eligible for the intervention in the following two years is expected to be low
- children in the control cohort would not be attending the intervention classes, which are aimed at Reception year pupils.

As a result, the benefits of a wait-list approach (mainly improving school recruitment and retention) were judged to outweigh the risks.

Timetable

Schools were recruited by local delivery partners between May 2016 and September 2017 (see Table 1). The online testing for baseline data collection was then conducted by schools between October and November 2016.

Following delays to baseline data collection, and in order to maximise the available time for parent recruitment in treatment schools that had completed baseline assessments, randomisation was conducted by NatCen in four waves over three weeks at the end of November and beginning of December 2016.

Families were recruited to the Family Skills intervention between November 2016 and January 2017, with additional recruitment and retention activities continuing throughout delivery. The intervention was delivered between January and March 2017. Post-intervention outcome data collection will take place in June - July 2017.

Table 1: Family Skills timetable

	2016	2017

	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Recruitment														
Baseline data collection														
Randomisation														
Intervention delivery														
Outcome data collection														

3. Protocol changes

If any changes to the protocol impact on the SAP, these should be specified here.

Changes to the protocol impacting on the statistical analysis plan are outlined below and referenced in more detail in relevant sections of the statistical analysis plan.

- The target number of schools to be recruited was increased to 155 and the window for school recruitment extended until the 9th of September 2016 due to challenges in recruitment of the target number of schools within the initially agreed recruitment window and concerns over potentially lower programme take-up than originally anticipated. This change has resulted in a more condensed timeline for the set-up and conduct of baseline data collection in schools.
- The eligibility criteria for schools have been revised to any school with a minimum of 6 pupils with EAL (including single form entry schools) due to challenges in recruiting a sufficient number of schools meeting the original inclusion criteria.
- The baseline literacy assessment window was extended until the end of November following challenges faced by schools in setting-up and completing baseline data collection within the new condensed timelines and technical difficulties experienced by schools when conducting baseline testing,
- Schools that have conducted baseline assessments for some but not all students were included in the trial and randomised to maximise the sample size available for the trial. Randomisation was conducted in four waves over three weeks at the end of November and beginning of December to maximise the available time for parent recruitment in treatment schools that have completed baseline assessments. More details available in section 4 below.
- The power calculations in this statistical analysis plan have been revised to reflect the sample sizes of the study sample at randomisation and most recent information about programme take-up. See section 5 for more detail.
- A post-intervention parent survey was intended allow the estimation of changes in the Home Literacy Environment (HLE) from a baseline measure (collected through a baseline parent survey). Improvements to the HLE are seen as a mechanism transmitting part of the causal force of Family Skills and acting on the primary attainment outcome. However, the baseline survey had a low response rate and high item non-response. Combined with the currently low take-up of the programme, it is unlikely that a post-intervention survey would be able to identify any impact on the HLE. As a result, in consultation with the EEF, we decided to cancel the post-intervention survey. As a result, the originally envisaged secondary outcome analysis will not be conducted.

4. Randomisation

The randomisation was carried out independently by NatCen Social Research at the end of November 2016 once baseline data collection has completed. Schools were randomly assigned to *Family Skills* or the control condition using a stratified randomisation, with the sample stratified by randomisation wave and delivery partner as outlined in more detail below.

Due to delays in schools completing baseline testing, the project team had to grapple with the need to ensure a large enough sample size of schools for a sufficiently well-powered trial and the competing need to initiate parent recruitment for the programme in schools assigned to the treatment condition. As a result, following agreement between the EEF, the developers and NatCen, the randomisation was conducted in 4 consecutive waves as outline below, with the wave acting as the primary stratification factor, and each wave being further stratified by delivery partner[8].

- Randomisation Wave 1 included 105 schools that have completed baseline attainment within the final agreed testing window closing on the 18th of November[9]. Randomisation was conducted on the 22nd of November 2016, two working days following the completion of baseline testing.
- Randomisation Wave 2 included seven schools that have completed baseline testing within the subsequent week, with randomisation taking place on the 29th of November 2017.
- Randomisation Wave 3 included one school that has completed baseline testing shortly after Wave 2 schools, with randomisation taking place at the end of the same day as Wave 2 (the 29th of November 2017)[10].
- Randomisation Wave 4 included two schools that completed baseline testing by the 6th of December, with randomisation taking place on the 7th of December 2017.

In each randomisation wave, the same randomisation procedure was used. The sample was stratified by delivery partner. This was done in order to ensure each provider has an equal number of treatment and control schools among those it has recruited and to maximise balance on delivery partner characteristics across the trial arms at randomisation.

The randomisation process was conducted in Stata as follows: ahead of random assignment, participating schools were stratified by a variable indicating the delivery partner that recruited the school to the study. Each school was then allocated a random number using a random number generator in Stata. Within each stratum, schools were arranged in descending order on the basis of their allotted random number. Two

[8] The original intention was to conduct just one wave of randomisation and stratify on delivery partner to ensure balance on delivery partner characteristics across the trial arms at randomisation. Stratifying by randomisation wave in the first instance meant the potential for some amount of treatment imbalance on delivery partner characteristics across the trial arms. However, the EEF the project team agreed a priori that this was an acceptable trade-off for improved trial power and a longer parent recruitment period for most schools

[9] The baseline testing window was extended from a two-week window starting on the 10th of October 2016 and ending on the 21st of October to a six weeks ending on the 18th of November due to schools requiring additional time to complete baseline testing.

[10] This school was randomised separately from Wave 2 schools due to uncertainty about when the school might complete baseline assessment at the time of Wave 2 randomisation and an urgency for delivery partners to initiate parent recruitment in schools that have completed baseline testing as soon as possible. The school was subsequently randomised on the same day when the NatCen evaluation team has been informed about the school having completed baseline testing. The school in Wave 3 was the only school among the Wave 2 and 3 schools within that delivery partner strata so the separate randomisation process for this school will not have affected treatment allocation balance across the sample.

groups of schools were formed within each stratum through assigning the first half of the schools in the arrangement to treatment group and the second to control group. For strata (delivery partners) with an odd number of schools the last school was randomly allocated to either treatment or control condition using the last replication correction procedure to minimise imbalance in treatment allocation across the sample. This was repeated across all strata[11].

The randomisation was carried out by an independent analyst within the evaluation team who was blinded with respect to the identity of the schools at treatment allocation and will remain blinded over the duration of the trial to prevent risk of bias during the statistical analysis.

The randomisation process has resulted in 59 schools randomly assigned to Group 1 (treatment group) and 56 schools to Group 2 (control schools). Table 2 below outlines the distribution of treatment and control schools at randomisation by delivery partners. As can be seen in the table school allocation to treatment and control was balanced across delivery partner strata at randomisation.

Table 2: Results of randomisation

Delivery partner	**Control**	**Treatment**	**Total**
Aspire Sussex Ltd (Formerly West Sussex Adult and Community Learning Service)	6	6	12
Buckinghamshire Family Learning	2	3	5
Doncaster Metropolitan Borough Council, Skills, Enterprise, Policy & Improvement	4	4	8
Ealing Adult Learning	1	1	2
Lancashire Adult Learning College	2	3	5
Learning Unlimited	5	4	9
Leicestershire Adult Learning Service	4	4	8
London Connected Learning Centre	5	4	9
Northamptonshire County Council, Adult Learning Service	6	6	12
Nottingham City Council	2	3	5
Nottinghamshire Community Learning and Skills Service (CLaSS)	3	3	6
Portsmouth City Council	4	5	9
Redbridge Institute	2	2	4
Renaisi	3	4	7
Sheffield City Council – Lifelong Learning, Skills and Communities	2	3	5
Smartlyte Ltd	5	4	9
Total	**56**	**59**	**115**

[11] The randomisation process used diverges slightly from that proposed in the trial protocol, including in the use of Stata rather than Excel. We have opted for these changes to ensure greater transparency and facilitate a more thorough peer-review of the randomisation steps taken, outlined in the Stata do file and available on demand.

5. Sample size calculations

The sample sizes and the associated minimum detectable effect sizes (MDES), expressed in standard deviations, are displayed in Table 1. The sample size calculations are focused on the main hypothesis, comparing the literacy attainment of EAL pupils in the Family Skills treatment schools to EAL pupils in the control schools.

The sample size calculations use a cluster-randomisation approach where the estimation equation or the MDES is a four level model with randomization at Level 3 (schools), and level 4 represents 16 stratification blocks[12]. The calculations assume:

- 80% statistical power (or a type II statistical error rate of 20%)
- a statistical significance level of 95% for a two-tailed test (or type I statistical error rate of 5%)
- an intraclass correlation coefficient[13] (ICC) at school level (i.e. the degree to which intervention effectiveness clusters within schools) of 0.11[14].
- an ICC at class level (ρ_2) (i.e. the degree to which intervention effectiveness clusters within classes) of 0.05
- 100% compliance with treatment assignment at the school and class-level

The calculations further are based on the total numbers of participating delivery partners, schools, classes and pupils and associated harmonic means[15] derived from baseline data available from randomised schools in autumn 2016 as outlined in Table 3 below.

Table 3: Sample sizes and harmonic means used in power calculations

	Total sample size	Harmonic mean	Arithmetic mean
Delivery partners	16		
Schools	115	5.88 / delivery partner	7.19 / delivery partner
Classes	251	1.84 / school	2.20 / school
Pupils	2499	6.39 / class	10.41 / class

Adjusting analysis for predictive baseline covariates reduces heterogeneity in outcomes and thus increases power. The proportion of variance explained can be high for educational interventions if pre-test scores are used in adjusted analysis[16]. We assumed the proportion of individual-level variance explained by individual level baseline attainment to be 0.54 based on values supplied by a co-author of a recent study of the impact of 27 school-based family literacy programmes on young children's

---

[12] Nianbo Dong and Rebecca Maynard, 'PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies', *Journal of Research on Educational Effectiveness* 6, no. 1 (1 January 2013): 24–67, doi:10.1080/19345747.2012.673143.

[13] The intraclass correlation coefficient is the between-cluster variance divided by the total variance

[14] Estimates based on the ICCs provided by the Education Endowment Foundation, calculated using data from the NPD 2013-4 academic year. Education Endowment Foundation, *Intra-Cluster Correlation Coefficients* (London: Education Endowment Foundation, 2015), https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/ICC_2015.pdf.

[15] The harmonic mean is a type of average and is recommended for use in power calculations to estimate minimum detectable effects when cluster sizes vary as it is more robust to extremely large outliers and therefore more conservative than other types of means (Dong and Maynard (2013), *op. cit.*). Where cluster sizes are identical, a harmonic mean equals an arithmetic mean.

[16] Howard S. Bloom, Lashawn Richburg-Hayes, and Alison Rebeck Black, 'Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions', *Educational Evaluation and Policy Analysis* 29, no. 1 (3 January 2007): 30–59, doi:10.3102/0162373707299550.

progress in reading and writing by the UCL Institute of Education[17]. Bloom estimates that the extent of school level variance explained by individual level baseline measures (R-squared at school level) ranges from 0.18 to 0.73 for covariates measuring attainment years prior to the intervention[18]. Below, we present the MDES assuming an R-squared at level one[19] of 0.54, and different assumed values of R-squared at school level[20,21].

As explained in Section 2, the study sample at level 1 will comprise all pupils identified by participating schools as EAL and whose parents/carers do not opt out of the trial prior to randomisation and for whom baseline measures are obtained. This is the study sample upon which intention to treat analysis will be performed. At the pupil or family level, however, we cannot assume that all pupils assigned to the intervention go on to participate. Thus the intention to treat sample will include a proportion of pupils and parents who do not take up Family Skills even though they are eligible for it and have not opted-out. This non-compliance or failure to take up a place on the programme will dilute the treatment effect and lower sample power. We have modelled the effect of non-compliance among eligible parents/carers and pupils on MDESs based on an approach suggested by Duflo, Glennerster, & Kremer (2006) and adapted for the particular CRT design we are proposing[22]. Assuming that delivery partners, schools and classes do not drop out of the study, we have considered the effects of 100%, 70% and 50% per cent average compliance rates among parents/carers and their children on the MDES reported in Table 4 columns 3-5.

Table 4 below presents a range of MDES for different assumptions about the predictive power of covariates and take up of the intervention by families. The MDES ranges from 0.19 in the most optimistic scenario (100% programme take up and 60% of variance explained by covariates) to 0.42 (30% programme take up and no variance explained by covariates). A recent update from the developers indicate a take-up/attendance rate of approximately 30 per cent (measured half-way through the programme delivery). Assuming an *R-squared at school level* of 0.2 as conservative but realistic scenario, we expect the trial to be able to detect an MDES of 0.41 standard deviations.

Table 4: Minimum detectable effect sizes: Intention-to-treat analysis

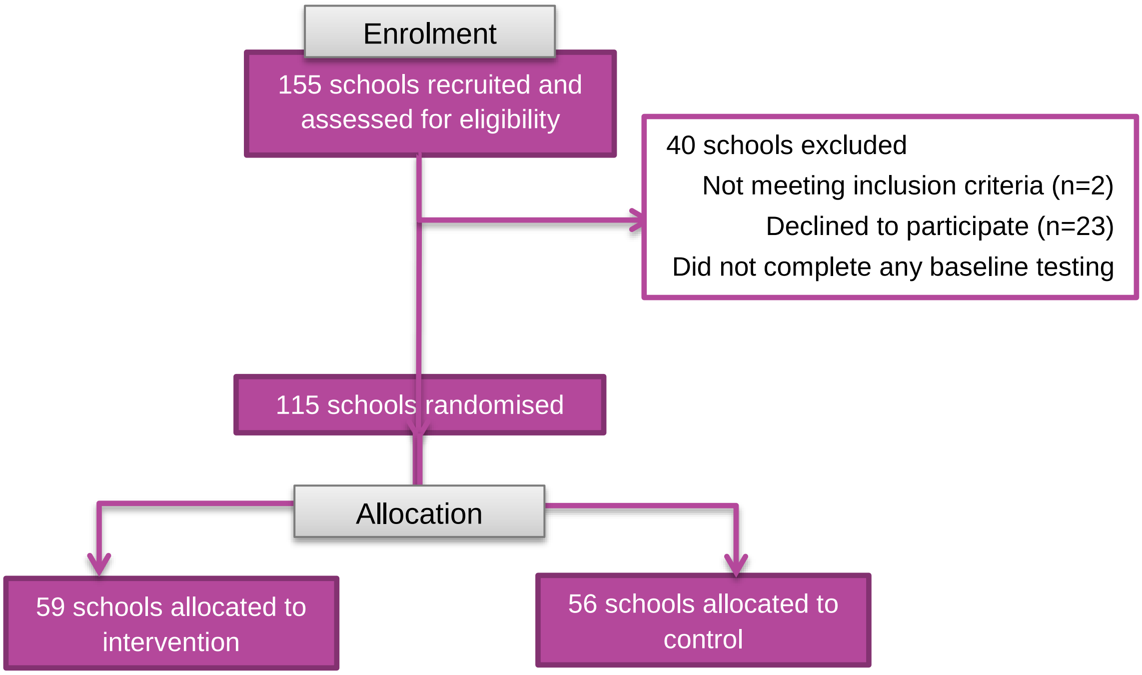
		Percentage programme take up at family level			
		100%	**70%**	**50%**	**30%**
Proportion of school-level variance explained by covariates	**0.0**	0.24	0.27	0.30	0.42
	0.2	0.23	0.25	0.29	0.41
	0.4	0.21	0.24	0.28	0.40
	0.6	0.19	0.22	0.27	0.39

[17] Jon Swain et al., 'The Impact of Family Literacy Programmes on Children's Literacy Skills and the Home Literacy Environment' (London: National Research and Development Centre for adult literacy and numeracy, 2015), http://www.nrdc.org.uk/wp-content/uploads/2015/11/Nuffield-Family-Literacy-Report.pdf.
[18] 'Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions'.
[19] Proportion of individual-level variance explained by baseline literacy attainment of reception year pupils participating in the trial.
[20] Proportion of school-level variance explained by the stratification variable (delivery partner) and school-level aggregate baseline literacy attainment of pupils participating in the trial.
[21] R-squared at class level was assumed to be zero as we do not propose to include class-level covariates in the primary analysis model.
[22] Esther Duflo, Rachel Glennerster, and Michael Kremer, 'Using Randomization in Development Economics Research: A Toolkit', Technical Working Paper (Cambridge, MA: NBER, December 2006), http://economics.mit.edu/files/806.

6. Follow up

The flow of participating schools through the initial stages of the trial using a CONSORT flow-diagram is briefly outlined below (Figure1). The diagram will be updated in the study report to present a complete summary of the flow of trial pupils and their schools from recruitment through baseline assessment, randomisation, post intervention assessment and analysis. The number of pupils and schools included or excluded at each stage will be clearly stated and the reasons for exclusion will also be stated.

Figure1: CONSORT Flow diagram



7. Data sources

The analysis will use data from five sources, as set out in Table 5:

Table 5: Data sources

Source	Data
School baseline and post-intervention surveys	• **Data on pupils participating in the trial** including name, Unique Pupil Number (UPN), Gender, Date of birth, Home postcode, Proficiency in English level, and Eligibilityfor Pupils Premium • **School level information such as** school type and number of reception year pupils and pupils with EAL[23] • **Data on other planned/conducted school literacy and EAL activities** for

[23] Note that from September 2016 EAL is no longer reported as a binary measure but on a five-point scale dividing pupils into “New to English”, “Early Acquisition”, “Developing Competence”, “Competent” and “Fluent”.

	reception year pupils and/or parents – collected at baseline and post-intervention surveys • **Data on school-level programme costs** (including staff time) – to be collected in the post-intervention survey
Pupil test	• CEM BASE Inspection Ready - reception year **baseline and end of year assessment of literacy**, numeracy and communication skills.
Registers	• **Data on treatment group parent/carer participating in the Family Skills programme**, including gender, relation to the pupil and residency status • **Information on take-up, drop-out and attendance (dosage)** of each treatment group parent/carer and pupil participating in the programme
Edubase	• **School level information:** • Establishment number as assigned by the DfE (LAESTAB, and Unique Reference Number (URN)) matched to participating school details provided by delivery partners

8. Outcome measures

Primary outcome

The primary outcome measure will be the raw results of the literacy attainment test CEM BASE Inspection Ready[24], obtained at follow-up among the EAL pupil sample. This test is an online literacy and numeracy assessment, which will be administered by Teaching Assistants or another member of school staff within the schools.

9. Analysis

Primary Intention to Treat (ITT) analysis

The primary analysis will explore the following hypothesis:

- **H1**: Eligible reception-year pupils whose families have been assigned to receive an offer to participate in the Family Skills programme will have better literacy outcomes on average than eligible pupils that have been assigned to the control group and

[24] Centre for Evaluation and Monitoring, 'BASE Progress and BASE Inspection Ready: Reporting & Feedback', 2015, http://www.cem.org/reporting-feedback.

who have been offered business-as-usual support for their literacy skills development.

The primary analysis of hypothesis H1 will be conducted on an intention-to-treat basis and will include all trial pupils pre-designated to receive treatment (i.e. those deemed EAL pupils by schools) subject to randomisation status, and for whom outcome data has successfully been collected. Following EEF guidance[25], evidence of effectiveness and reported effect sizes will be obtained from a baseline-adjusted analysis, in which the dependent variable is the result of the CEM test, and effects are estimated through a multilevel linear model containing a dummy variable indicator capturing treatment/control group membership, the stratification variables (delivery partner), and pupil level baseline test scores at individual and aggregated at school level. The model analysed will be a four level model: pupil (level 1), class (level 2), school (level 3) and delivery partner (level 4); with levels 2 and 3 modelled as random effects (using a random intercept model) and the fourth as fixed effects (see Formula (1)).

$$\begin{aligned} Literacy_{ijk} = \beta_0 + \beta_1 baseline_{ijk} + \beta_2 intervention_k + \beta_3 partner_k \\ + \beta_4 baselineaggregate_k + w_k + u_{jk} + e_{ijk} \end{aligned} \quad (1)$$

where *i* presents the individual, *j* presents the class, *k* presents the school, and w_k and u_{jk} are the school and class random effects, respectively. The intervention effect is estimated by β_2.

The analysis will run in ~~R~~ *Stata 14.1* ~~in the latest available version~~, and analysts will not be blind to treatment status of the groups.

Imbalance at baseline

Baseline characteristics will be summarised for trial pupils by treatment and control group across schools and pupils. Variables available at pupil level will be presented at that level, otherwise at school level.

Continuous variables will be summarised with descriptive statistics (n, mean, standard deviation, range and median).

At school level, the comparison will cover:

- School type
- English as additional language (EAL) ~~status~~composition
- Special Educational Needs (SEN) ~~status~~composition

At the pupil level, the following baseline comparisons will be presented:

- Pupil premium eligibility
- Gender
- Birth term
- English Language Fluency level (assessed using Proficiency in English (PiE) rating)

[25] Education Endowment Foundation, *Policy on Analysis for EEF Evaluations* (London: Education Endowment Foundation, 2015), https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_for_EEF_evaluations.pdf.

- Baseline CEM literacy attainment

Imbalance on baseline covariates between the treatment and control groups in the sample as analysed will be assesses for the covariates listed above using the appropriate statistical test (two independent sample *t*-test for continuous variables and Fisher's exact test for categorical variables) with a p-value of 0.05 or smaller considered as indication of covariate imbalance. Differences in baseline tests will be reported as effect size.

Sensitivity analyses

A range of sensitivity analyses will be carried out to explore the robustness of the main finding. The following analyses will be carried out:

- An unadjusted analysis that does not include the baseline covariates;
- Including a wider range of prognostic covariates to increase power: Including reception-level proportions of pupils with SEN and disability, school type, individual-level indicators of Pupil Premium eligibility, gender and date of birth, and if suitable Proficiency in English level as additional co-variates;
- If baseline imbalance is found, an analysis adjusting for baseline imbalance will be carried out.

Missing data

For the main primary analysis of pupil outcomes, we expect to lose around 10% of pupils due to moves and other external factors influencing participation in the final outcome testing. Given that our pupil sample is restricted to pupils who have been tested at baseline and we do not include any other co-variates in the main analysis, we do not expect further losses due to lack of covariate data. We anticipate no attrition at school level.

For the main secondary analysis, we may experience a larger loss due to low and potentially differential response rates between the treatment and control group to the follow-up parent survey.

For the main primary and secondary analyses, we will assume that missing data are missing completely at random and use complete case analysis. We will then conduct sensitivity analyses to assess the robustness of the inferences about treatment effects to alternative assumptions about the mechanisms leading to missing data.

We will explore the number, pattern and likely reasons for missing outcome data for both the primary and secondary outcomes. We will report the number of pupils/parents with missing outcome data by treatment arm and present comparisons of pupil/parent baseline characteristics with observed and missing outcomes using cross-tabulations. We will then run a drop-out model to assess whether any existing covariates predict the observed loss-to-follow up pattern. We will estimate the propensity score using multivariate logistic regression. In this case, the propensity score can be thought of as the conditional probability that a pupil has been tested or a parent responded in the post-intervention questionnaire given the set of existing covariates capturing pupil/parent characteristics.

If loss-to-follow up can be predicted using existing covariates, we will conduct sensitivity analyses under the assumption that outcome data are missing at random. We will use multiple imputation to infer the likely results of those lost to follow-up. For each outcome measure, we will impute a number of new datasets[26] using a four level linear model: pupil (level 1), class (level 2), school (level 3) and delivery partner (level 4); with 3 modelled as random effects (using a random intercept model) and level 2 and 4 as fixed effects. The model will include all variables in the main analysis (baseline outcome measure and stratification variables); any variables predictive of missingness; and variables associated with the outcome, i.e. reception-level proportions of pupils with SEN and disability, school type, individual-level indicators of Pupil Premium eligibility, gender and date of birth, and Proficiency in English. The estimated regression equations will then be used to generate predicted values for the cases with missing data and estimate treatment effects and standard errors under the alternative assumption.

We will then conduct sensitivity analysis to explore the robustness of the missing at random assumption. We will use simple imputation methods to assess the robustness of the inferences about treatment effect to extreme case scenarios of data missing not at random. We will impute the highest possible scores for missing outcome values in the treatment group and lowest possible scores for the missing values in the control group to model the best case scenario, and vice versa for the worst case scenario, and re-estimate the treatment effects in both cases. This analysis will provide a range of treatment effects under the most extreme cases of data missing not at random and therefore the upper and lower bounds of the likely effect sizes under the assumption that data is not missing at random. However, these estimates will need to be interpreted with caution as this method does not adjust for the fact that the imputation process involves uncertainty about the missing values and is therefore likely to produce standard errors that are underestimated.

On-treatment analysis

The on-treatment analysis will explore the following hypothesis:

- **H2**: Pupils whose families participate in *Family Skills* will have a better literacy outcome than pupils of non-participating eligible families.

To assess the impact of the Family Skills programme on those that have taken up the treatment, we will estimate a complier average causal effect (CACE). Given that trial pupils and parents in control schools do not have access to the Family Skills programme, the CACE can be estimated under the assumption of one-sided non-compliance. We will label any trial pupils in the treatment group whose parents have not taken up the programme (i.e., have not attended any sessions) as non-compliant and estimate a complier average treatment effect (CACE) by dividing the ITT estimate by the share of compliers.

$$CACE = \frac{ITT_y}{\Pr(Compliers)} \tag{2}$$

Assuming that random assignment to treatment does not affect the outcomes of individuals who do not comply with treatment (an assumption known as the exclusion restriction), this effect estimate provides an unbiased estimate of the average impact of

[26] The number of imputed datasets will be selected based on guidance by Graham et al. (2007) who approached the problem in terms of loss of power for hypothesis testing. Based on simulations (and a willingness to tolerate up to a 1 percent loss of power), they recommended 20 imputations for 10% to 30% missing information, and 40 imputations for 50% missing information. Source: John W. Graham, Allison E. Olchowski, and Tamika D. Gilreath, 'How Many Imputations Are Really Needed? Some Practical Clarifications of Multiple Imputation Theory', *Prevention Science* 8, no. 3 (1 September 2007): 206–13, doi:10.1007/s11121-007-0070-9.

the programme on the treated (ATT), where those treated in this case are defined as those that have taken up the programme and attended at least one session.

We will estimate a complier average causal effect for both the primary and secondary outcome, using the ITT effect estimates from the main analyses (adjusting for the baseline outcome and stratification variables only) in each case.

Additional analyses

An additional analysis will explore hypothesis 5:

- **H5:** The number of Family Skills sessions attended by the parent/carer will be positively correlated with participating pupils' literacy outcomes, controlling for other factors.

We will explore the variation in primary outcomes across the treated sample by parent participation rates in the Family Skills programme using random effects/ multilevel multiple linear regression. This analysis will be conducted exclusively on the sub-sample of treatment-group pupils and parents that have taken up the programme. We will construct measures of dosage for those attending at least one session from the registers co-developed by NatCen and the developers and collected by session instructors. The findings of this analysis will not be able to identify a causal effect between the number of sessions attended and outcomes due to the non-experimental nature of the comparison. However, the findings can provide useful indicative evidence of the relationship between participation and outcomes that may warrant further research using more robust designs.

The analysis will be conducted using a three-level linear regression model containing an indicator capturing "dosage" as the independent variable of interest, random effects at the level of the school and fixed effects at the level of the delivery partner, with the addition of a range of prognostic covariates controlling for other factors affecting pupil outcomes. In a change to what was stated in the protocol, these will include pupil level baseline test scores at individual- and school-level reception-level proportions of pupils with SEN and disability, school type, individual-level indicators of Pupil Premium eligibility, pupil gender, date of birth, and if suitable Proficiency in English level[27] collected through the baseline school information form and parent relation to child, gender and residency status collected through the Family Skills programme registers.

Subgroup analyses

The sub-group analyses will explore the following hypotheses:

- **H3**: *Family Skills* will have a different (higher or lower) impact on pupils participating in the intervention and eligible for pupil premium.
- **H4**: *Family Skills* will have a different (higher or lower) impact on female and male pupils participating in the intervention.
- **H5**: *Family Skills* will have a different (higher or lower) impact on pupils with different baseline English language skills (assessed using Proficiency in English

[27] However, note that schools had the option to mark children as "not yet assessed" in the September 2016 school census, so we may encounter a high proportion of missing values for this variable at baseline. If the proportion of missing values exceeds 80 per cent, we propose to exclude this variable from the analysis to avoid the associated loss of precision in the estimate.

(PiE) rating[28] if less than 20% of PiE data is missing, or baseline CEM literacy attainment otherwise).

Subgroup impacts on the primary outcome will be estimated for groups defined by pupil premium status, and gender and English language proficiency. We will estimate sub-group impacts by Proficiency in English rating at baseline, if at least 80 per cent of pupils at baseline have been rated using the 5-point Proficiency in English scale in the School Census (and these ratings have been reported by schools in the baseline information form). If the number of assessed pupils at baseline is deemed insufficient, we will estimate sub-group impact for groups defined by above and below-median baseline literacy attainment on the CEM BASE assessment.

Estimation of subgroup effects on the primary outcome will involve the re-estimation of the adjusted model described in the previous paragraph with the addition of a further covariate for the particular subgroup concerned. This additional covariate will be interacted with the treatment/control group indicator. Where the coefficients resulting from this interaction reach statistical significance at the 95 per cent level, separate models will be estimated and reported for each subgroup.

Effect size calculation

The calculation of effect sizes will follow the methodology for Hedges' *g* with a three-level cluster-randomised trial and unequal sample sizes; scaling the mean differences between treatment and control by the total standard deviation within-treatment groups[29]. The formulas given by Hedges are amended, following Borenstein[30], to use covariate-adjusted estimates. Accordingly, the difference in adjusted means is scaled by the pooled sample variance of the post-test measures (i.e. the unadjusted variance).

We will use the formulae outlined below for the calculation of effect sizes and 95% confidence intervals.

The point estimate of the effect size *g* will be calculated as the difference between adjusted treatment groups means, scaled by the within-treatment standard deviation and adjusted to take account of clustering at school and class-level, as follows:

$$g_{WT} = J \times (\frac{\overline{Y}_{adj}^{T} - \overline{Y}_{adj}^{C}}{S_{WT}})\sqrt{1 - \frac{2(p_U - 1)\rho_S + 2(n_U - 1)\rho_C}{N - 2}}, \tag{3}$$

where *J* is Hedges' bias correction to Cohen's *d*, given by:

$$J = 1 - \left(\frac{3}{4(n_1 + n_2) - 9}\right). \tag{4}$$

The standard deviation is the square root of the estimated pooled population variance S_{WT}^2, calculated as:

[28] From September 2016, schools are required to record a Proficiency in English rating (using the 5-point Proficiency in English scale) in the School Census for all pupils' in reception year and above for whom primary language is anything other than 'English' or 'Believed to be English'. Schools with limited capacity to fully compete the assessment in September 2016 can classify pupils as 'Not yet assessed' in the first instance, with all schools expected to complete the assessment for all pupils by the Spring 2017 census.

[29] Larry V. Hedges, 'Effect Sizes in Three-Level Cluster-Randomized Experiments', *Journal of Educational and Behavioral Statistics* 36, no. 3 (1 June 2011): 360, doi:10.3102/1076998610376617, equation 31 and following.

[30] Michael Borenstein, 'Effect Sizes for Continuous Data', in *The Handbook of Research Synthesis and Meta-Analysis*, ed. Harris M. Cooper, Larry V. Hedges, and Jeffrey C. Valentine, 2nd edition (New York: NY: Russell Sage Foundation, 2009), 221–36.

$$S_{WT}^2 = \frac{\sum_{i=1}^{m^T}\sum_{j=1}^{p_i^T}\sum_{k=1}^{n_{ij}^T}(Y_{ijk}^T - \overline{Y}^T)^2 + \sum_{i=1}^{m^T}\sum_{j=1}^{p_i^T}\sum_{k=1}^{n_{ij}^T}(Y_{ijk}^C - \overline{Y}^C)^2}{N-2}. \tag{5}$$

The school-level intraclass correlation ρ_S is defined by:

$$\rho_S = \frac{\sigma_{BS}^2}{\sigma_{BS}^2 + \sigma_{BC}^2 + \sigma_{WC}^2} = \frac{\sigma_{BS}^2}{\sigma_{WT}^2} \tag{6}$$

and the class-level intraclass correlation $\rho_C$by:

$$\rho_C = \frac{\sigma_{BC}^2}{\sigma_{BS}^2 + \sigma_{BC}^2 + \sigma_{WC}^2} = \frac{\sigma_{BC}^2}{\sigma_{WT}^2} \tag{7}$$

Other terms are calculated as follows:

$$p_U = \frac{N^C \sum_{i=1}^{m^T}(\sum_{j=1}^{p_i^T} n_{ij}^T)^2}{NN^T} + \frac{N^T \sum_{i=1}^{m^C}(\sum_{j=1}^{p_i^C} n_{ij}^C)^2}{NN^C}, \tag{8}$$

$$n_U = \frac{N^C \sum_{i=1}^{m^T}\sum_{j=1}^{p_i^T}(n_{ij}^T)^2}{NN^T} + \frac{N^T \sum_{i=1}^{m^C}\sum_{j=1}^{p_i^C}(n_{ij}^C)^2}{NN^C}, \tag{9}$$

and

$$N = N^T + N^C = \sum_{i=1}^{m^T}\sum_{j=1}^{P_i^T} n_{ij}^T + \sum_{i=1}^{m^C}\sum_{j=1}^{P_i^C} n_{ij}^C. \tag{10}$$

We will calculate the 95% confidence interval as follows:

$$g_{WT} - 1.96v_g \le \delta_T \le g_{WT} + 1.96v_g \tag{11}$$

where v_g is the variance of the effect size estimator g_{WT}, conservatively approximated by:

$$v_{\{g_{WT}\}} \approx \frac{(1 + (p_U - 1)\rho_S + (n_U - 1)\rho_C)(1 - r^2)}{\tilde{N}} + \frac{d_{WT}^2}{2(M-2) - q}, \tag{12}$$

where

$$\tilde{N} = \frac{N^T N^C}{N^T + N^C}, \tag{13}$$

and q is the number of covariates included in the model.

10. References

Bloom, Howard S., Lashawn Richburg-Hayes, and Alison Rebeck Black. ‘Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions’. *Educational Evaluation and Policy Analysis* 29, no. 1 (3 January 2007): 30–59. doi:10.3102/0162373707299550.

Borenstein, Michael. ‘Effect Sizes for Continuous Data’. In *The Handbook of Research Synthesis and Meta-Analysis*, edited by Harris M. Cooper, Larry V. Hedges, and

Jeffrey C. Valentine, 2nd edition., 221–36. New York: NY: Russell Sage Foundation, 2009.

Centre for Evaluation and Monitoring. ‘BASE Progress and BASE Inspection Ready: Reporting & Feedback’, 2015. http://www.cem.org/reporting-feedback.

Dong, Nianbo, and Rebecca Maynard. ‘PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies’. *Journal of Research on Educational Effectiveness* 6, no. 1 (1 January 2013): 24–67. doi:10.1080/19345747.2012.673143.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. ‘Using Randomization in Development Economics Research: A Toolkit’. Technical Working Paper. Cambridge, MA: NBER, December 2006. http://economics.mit.edu/files/806.

Education Endowment Foundation. *Intra-Cluster Correlation Coefficients*. London: Education Endowment Foundation, 2015. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/ICC_2015.pdf.

———. *Policy on Analysis for EEF Evaluations*. London: Education Endowment Foundation, 2015. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_for_EEF_evaluations.pdf.

Graham, John W., Allison E. Olchowski, and Tamika D. Gilreath. ‘How Many Imputations Are Really Needed? Some Practical Clarifications of Multiple Imputation Theory’. *Prevention Science* 8, no. 3 (1 September 2007): 206–13. doi:10.1007/s11121-007-0070-9.

Hedges, Larry V. ‘Effect Sizes in Three-Level Cluster-Randomized Experiments’. *Journal of Educational and Behavioral Statistics* 36, no. 3 (1 June 2011): 346–80. doi:10.3102/1076998610376617.

Swain, Jon, Olga Cara, John Vorhaus, and Jenny Litster. ‘The Impact of Family Literacy Programmes on Children’s Literacy Skills and the Home Literacy Environment’. London: National Research and Development Centre for adult literacy and numeracy, 2015. http://www.nrdc.org.uk/wp-content/uploads/2015/11/Nuffield-Family-Literacy-Report.pdf.