Statistical Analysis Plan
EasyPeasy

Evaluator: Durham University
Principal investigator: Dr Lyn Robinson-Smith
(maternity cover Vic Menzies)

Education Endowment Foundation

PROJECT TITLE	Independent Evaluation of the EasyPeasy programme
DEVELOPER (INSTITUTION)	EasyPeasy
EVALUATOR (INSTITUTION)	Durham University
PRINCIPAL INVESTIGATOR(S)	Dr Lyn Robinson-Smith (maternity cover Vic Menzies)
TRIAL (CHIEF) STATISTICIAN	Caroline Fairhurst, York Trials Unit, University of York
SAP AUTHOR(S)	Caroline Fairhurst
TRIAL REGISTRATION NUMBER	ISRCTN22325174
EVALUATION PROTOCOL URL OR HYPERLINK	Please see amended protocol here.

SAP version history

VERSION	DATE	REASON FOR REVISION
1.0	20/07/2018	Original
1.1	16/08/2018	Ranges for scaled scores for CELF- Preschool 2 UK provided. Typo corrected (CSBQ is a 34, not 33, item questionnaire). Clarification of how CSBQ subscale scores calculated included.

This analysis plan was written post-randomisation and prior to receipt of any outcome data, and deals only with the statistical analysis of the impact evaluation.

Table of contents

Introduction

EasyPeasy is a smartphone programme for parents of preschool aged children. It is designed to improve early child development through increasing positive parent-child interactions and learning at home. EasyPeasy sends regular game ideas to parents that they can play with their children, combined with a text schedule which includes information on child development and explanation of each of the games. The design of the programme applies behavioural insights to help seed positive habits of play and interaction at home by sending tailored prompts, encouragement, and reminders to parents.

Parents receive EasyPeasy communications via text message (SMS). Parents receive a series of videos via SMS which demonstrate games (weekly) and separate text explanations (known as the 'text schedule' of each game) encouraging them to play with their children.

EasyPeasy also provides a function whereby the child's early years provider can track parental engagement with the programme and create digital communities for parents to communicate and share experiences of the EasyPeasy activities (Jelley *et al*. 2016).

In a previous evaluation, Jelley *et al*. (2016) reported that EasyPeasy led to moderate positive effects on parenting self-efficacy and on children's cognitive self-regulation (parental reports), an increase in parental consistency with discipline and boundaries and improvements to child concentration and persistence (parental reports). EasyPeasy presents as a low-cost, accessible, non-intrusive intervention that does not place great demands on parents' time or resources. As the content is evidence-based, covering all areas of the Early Years curriculum, EasyPeasy shows potential as a quality learning resource to support parents and positively impact on the home learning environment and early years attainment.

Primarily the intervention will be evaluated with children in nursery classes attached to state schools in England. Eligible children will be aged 3+ years and due to enter Reception class in the academic year 2018/19. At least one early years teacher (the 'pod leader') will receive training from EasyPeasy to implement the programme within their setting and be responsible

for the recruitment of parents into the programme and for the setting's pod administration used to foster continued parental engagement.

Nurseries will be randomly allocated to receive the intervention in 2017-18, or to a 'wait-list' control to receive the intervention in 2018-19.

The trial aims to answer the following research questions:

1. What is the impact of the EasyPeasy intervention on the language and communication development of children aged 3-4 years old? [Primary Outcome]

2. What is the impact of the EasyPeasy intervention on the self-regulation of children aged 3-4 years old? [Secondary Outcome 1]

3. What is the impact of the EasyPeasy intervention on the social-emotional development of children aged 3-4 years old? [Secondary Outcome 2]

4. How effective is the EasyPeasy intervention at improving the home learning environment of children aged 3-4 years old? [Secondary Outcome 3]

Design overview

This is a pragmatic two armed cluster randomised controlled trial, with random allocation at the school level in a 1:1 ratio to:

▪ Intervention – schools allocated to receive the EasyPeasy intervention (a five-month intervention); or

▪ Control – schools allocated to continue with usual early years

Eligible settings were state funded schools whose child population includes children who are three years old. The recruitment of schools with an average ever-Free School Meal (FSM) percentage of ≥30% overall was desirable to ensure we were targeting those with the highest levels of deprivation. Schools with <30% FSMEVER and PVI nurseries were considered as a back–up if recruitment of schools with preferred characteristics was problematic. Eligible schools should not have previously been involved in the EasyPeasy intervention.

Parents of children who are three years old at the start of the intervention and due to start reception class in September 2018-19 were eligible to participate in the trial.

Data will be collected at baseline, prior to randomisation of the school (December 2017/January 2018) (or post-randomisation but before the start of the intervention for the HOME assessment, see section *Secondary outcomes*), and post-intervention (June/July 2018).

Trial type and number of arms	Two-armed cluster randomised controlled trial (random allocation at school level)
Unit of randomisation	School
Minimisation factor	Number of children with parental consent per school (2 levels; <14; ≥14)

Primary outcome	variable	Language and communication
	measure (instrument, scale)	Clinical Evaluation of Language Fundamentals Preschool 2 UK (CELF- Preschool 2 UK) - Core Language Score
Secondary outcome	variable	Language and communication
	measure (instrument, scale)	Clinical Evaluation of Language Fundamentals Preschool 2 UK (CELF- Preschool 2 UK) – individual subscale scores
Secondary outcome	variable	Self-regulation and social-emotional development
	measure (instrument, scale)	Child Social Behaviour Questionnaire (CSBQ) – subscales: Sociability, Externalising, Internalising, Prosocial, Behavioural Self-regulation, Cognitive Self-Regulation, Emotional Self-regulation
Secondary outcome	variable	Home environment
	measure (instrument, scale)	Home Observation and Measurement of Environment (HOME)

Outcome measures

Primary outcome

The primary outcome measure is the Clinical Evaluation of Language Fundamentals Preschool 2 UK (CELF- Preschool 2 UK) developed by Wiig, Secord and Semel[1], which provides a measure for expressive and receptive language skills in young children. The assessment is multiple choice and requires the child to simply point to the picture that is described by the test administrator, e.g. "point to the cat after I have pointed to the monkey"; there are multiple pictures for the child to choose from, the correct picture (cat) and other distractor pictures. Each subtest of the measure includes between 20 and 24 items, and discontinuation rules. For the purposes of this trial, the following norm-referenced subtests of the measure will be administered to all children with parental consent pre-randomisation and post-intervention (for each subtest, a raw score and scaled score will be calculated):

- Sentence Structure (22 Items. Discontinue rule: 5 consecutive zero scores) – raw score from 0 to 22, with a higher score indicating a better outcome (scaled scores range 1-18).
- Word Structure (24 Items. Discontinue rule: 8 consecutive zero scores) – raw score from 0 to 24, with a higher score indicating a better outcome (scaled scores range 1-19).
- Expressive Vocabulary (20 items. Discontinue rule: 7 consecutive zero scores) – raw score from 0 to 40, with a higher score indicating a better outcome (scaled scores range 1-19).

---

[1] See: http://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildLanguage/CELF-Preschool2UK/CELF-Preschool2UK.aspx

- Concepts and Following Directions (22 Items. Discontinue rule: 6 consecutive zero scores) – raw score from 0 to 22, with a higher score indicating a better outcome (scaled scores range 1-14).

A Core Language Standard Score will be derived by totalling the scaled subtest scores from the Sentence Structure, Word Structure and Expressive Vocabulary subtests. This score will serve as the primary outcome. The individual subtest scores will serve as secondary endpoints.

Independent test administrators (e.g. researchers, agency supply teachers) will conduct the assessment pre- (before randomisation) and post-intervention. Test administrators will be blinded to the school's random allocation for post-intervention testing.

Secondary outcomes

CLINICAL EVALUATION OF LANGUAGE FUNDAMENTALS PRESCHOOL 2 UK

The scaled individual subtest scores of the CELF- Preschool 2 UK will serve as secondary outcomes as described above.

CHILD SELF-REGULATION AND BEHAVIOUR QUESTIONNAIRE (CSBQ)

Self-regulation and behaviour scores gathered using the Child Self-regulation and Behaviour Questionnaire (CSBQ) (Howard & Melhuish, 2016). The CSBQ is a 34-item questionnaire pertaining to children's everyday behaviours related to children's social and emotional development and self-regulation (e.g., persists with difficult tasks and waits their turn in activities). It takes approximately 5 minutes to complete. CSBQ measures include:

- Sociability: Items 1, 4, 9, 16 (reversed), 22 (reversed), 27, 32
- Externalising: Items 3, 20, 23, 26, 28
- Internalising: Items 17, 21, 25, 33, 34
- Prosocial behaviour: Items 15, 19, 24, 27, 30
- Behavioural Self-Regulation: Items 7 (reversed), 13, 15, 29 (reversed), 30, 31 (reversed)
- Cognitive Self-Regulation: Items 5, 6, 8, 12, 18
- Emotional Self-Regulation: Items 2, 10, 11 (reversed), 14 (reversed), 23 (reversed), 26 (reversed)

For each item, the test administrator is asked to evaluate the child's frequency of target behaviours on a 5-point scale (1=not true – 5=very true). All subscales contain at least 5 items. Subscale scores are calculated by taking the average of the component item scores (sample score sheet available at http://www.eytoolbox.com.au/toolbox-data). The assessment will be completed by the children's class teachers for all children with opt-in parental consent pre- (prior to randomisation) and post-intervention; therefore, the outcome assessment will not be conducted blind to group allocation.

HOME OBSERVATION MEASURE OF THE ENVIRONMENT

The Home Observation Measure of the Environment (HOME) is intended for use by practitioners and researchers and is a descriptive profile which yields a systematic assessment of a child's home environment to measure, within a naturalistic context, the

quality and quantity of stimulation and support available to the child through measuring the active receipt of inputs from objects, events and transactions occurring within the home environment (Bradley 1993). The 'Early Childhood HOME' is suitable for use among 3-6 year olds and involves a visit by a researcher to the home, and includes an interview with the main caregiver whilst the child is present and awake. It is made up of 55 items that are grouped in eight different subscales that are scored in a binary manner (Yes/No). The 6 (of 8[2]) subscales relevant to the learning aims of EasyPeasy, and so will be used here, are:

1. Learning materials
2. Language stimulation (between child and caregiver)
3. Responsivity (verbal interaction between child and caregiver)
4. Academic stimulation
5. Modelling
6. Variety of activities and parental interaction

The instruction manual allows for totalling of subscales to give an overall score; however, this includes all 8 subscales. We are just using 6, but still believe it will be valid to calculate a total score with these 6 subscales; though will also consider the individual subscales also.

We are proposing that the HOME inventory be used in a sub-sample of approximately 50 households in both the intervention (n=25) and control groups (n=25), pre- and post-intervention, based on those that agree to the assessment at baseline. Pre-intervention HOME visits for both the intervention and control groups will be conducted in the weeks immediately following randomisation, prior to parents starting the intervention. Parents who have already provided consent for their child to participate in attainment measures as part of the evaluation will be invited to take part in the HOME visits. These visits will be conducted by the evaluator's appropriately trained Research Associate and a research assistant, who will be blind to group allocation. As incentive to parents to participate in the HOME visits, they will be provided with a £50 Love2Shop gift voucher after the final visit.

The use of the HOME inventory will allow us to measure the impact to learning that we anticipate EasyPeasy could have within home environments who receive the intervention. This measure will allow us to investigate how and why the intervention leads to improvements to the home learning environment, which can be determined through consideration of the different HOME subscales i.e. is it through changes to environment, better quality learning resources and/or more child parent interaction?
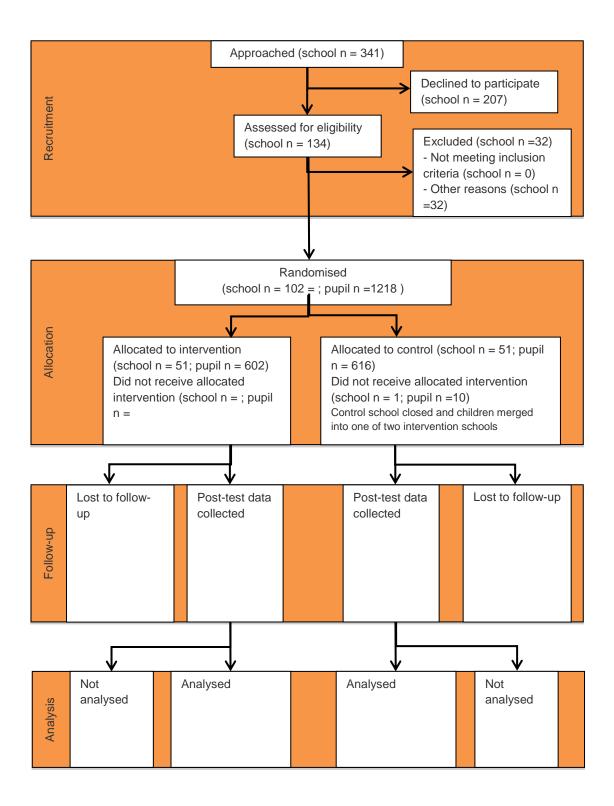
FUTURE OUTCOMES

Relevant data (child's name, date of birth, school destination, and Unique Pupil Number (UPN), where available) will be collected as part of the evaluation so that matching reference numbers can be provided for participating children to enable the long-term tracking of the EasyPeasy intervention through the National Pupil Database and, where available, the Early Years Foundation Stage Profile (EYFSP).

Follow-up

---

[2] Data will not be collected on the following HOME subscales: Physical environment and acceptance.

Recruitment	Approached (school n = 341)	
		Declined to participate (school n = 207)
	Assessed for eligibility (school n = 134)	
		Excluded (school n =32) - Not meeting inclusion criteria (school n = 0) - Other reasons (school n =32)

Randomised
(school n = 102 = ; pupil n =1218)

Allocation	Allocated to intervention (school n = 51; pupil n = 602) Did not receive allocated intervention (school n = ; pupil n =	Allocated to control (school n = 51; pupil n = 616) Did not receive allocated intervention (school n = 1; pupil n =10) Control school closed and children merged into one of two intervention schools

Follow-up	Lost to follow-up	Post-test data collected	Post-test data collected	Lost to follow-up

Analysis	Not analysed	Analysed	Analysed	Not analysed

Sample size calculations overview

The sample size calculations included here are those included in the amended protocol. Due to issues with recruitment (some schools being unable to recruit enough parents, or return required paperwork) and pre-testing causing some schools to withdraw, there was some concern about not reaching the number of schools specified in the original protocol.

Discussions were therefore held during the recruitment and pre-test period, between EEF, the EasyPeasy team and the evaluation team about the sample size calculations assumptions and expected effect size as well as what would be an acceptable number of

schools. The sample size calculations now included were completed by EEF (TE), using parameters discussed with the evaluation team. The sample size calculation was subsequently reviewed by, and agreed with, the trial statistician. We made the following assumptions for sample size estimate: an intra cluster correlation of 0.11 and 10 children per school whose parents have provided consent. We assumed the proportion of variance explained by covariates at the individual level (Level 1 – pupil level) to be 0.25 and at the cluster level to be 0.16 (Level 2 – setting level). Based on 102 schools (the number still in the trial at the time of discussions) (approx. 1020 children; 505 children per arm), we would have 80% power to show a difference of 0.224 of an effect size between the control and intervention groups.

At the piloting stage (conducted by Oxford University and funded by The Sutton Trust), the positive effect sizes of parents' self-efficacy regarding discipline and boundaries and child cognitive self-regulation were 0.51 and 0.44, respectively. We expect the effect size to be reduced for this study in line with Slavin and Smith (2009) who find that scaling to a larger sample is associated with two or three times smaller observed effect sizes.

In the original protocol we aimed to recruit 120 schools and 13 children per school to allow us to deal with a moderate amount of setting level attrition. However, as fewer schools were randomised (n = 102), we will continue to try to minimise attrition from the project in order to maintain all 102 schools at post-test. While the sample size calculations described here are presented in the amended, and current, version of the protocol (v2) the total number of pupils included at randomisation has been updated from the amended protocol following a subsequent review of assessment forms completed and parental consent.

The sample size calculations were conducted using MDES Calculator for Two-Level Cluster Random Assignment Design (CRA2_2) Treatment at Level 2 by the PowerUp tool (http://web.missouri.edu/~dongn/PowerUp.htm).

		Protocol	Randomisation
MDES		0.224	0.216
Pre-test/ post-test correlations	level 1 (pupil)	0.25	0.25
	level 2 (school)	0.16	0.16
Intracluster correlations (ICCs)	level 2 (school)	0.11	0.11
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		2	2
Average cluster size		10	11.94
Number of schools	intervention	51	51
	control	51	51
	total	102	102
Number of pupils	intervention	510	602
	control	510	616

	total	1020	1218

Randomisation

Randomisation was undertaken by the independent trial statistician, Caroline Fairhurst, who had no involvement in the recruitment of schools. Minimisation was used based on eligible cohort size as measured by the total number of children consented per school. A cut off of <14/≥14 was used as this was the median number of pupils consented per school in the first batch of schools randomised.

Randomisation was conducted in two batches due to delays in collecting pre-test data from schools. The first batch took place on 19th December 2017 and included 45 schools that had completed (or mostly completed) the CELF pre-test. Where some CELF assessment still needed to be completed, these schools were not informed of their randomisation until it was completed (8/45 schools). The second batch of randomisation took place on the 23rd January 2018 and included the remaining 57 schools that had completed (or mostly completed) the CELF assessment. One school still had some CELF assessments to be completed and they were informed of their allocation the following day after completion of the assessment.

Analysis

The statistical analysis proposed follows the most recent *revised EEF Statistical Analysis Guidance (2018)* available *here*.

Analysis will be conducted in Stata v15 (or later, to be confirmed in the final report) using the principles of intention to treat, where data are available, including all schools and pupils 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.

A CONSORT diagram will be produced to show the flow of schools and pupils through the trial.

Primary outcome analysis

The primary objective of this study is to investigate the effectiveness of the intervention on the language and communication skills of the participating children. Unadjusted raw and scaled outcome scores will be summarised by trial arm. The correlation between pre and post-intervention scores will be provided, for both the individual subscale scores and the total Core Language Standard Score. Histograms of pre- and post-test data distribution will also be presented. The difference in attainment between pupils in the intervention group and those in the control group will be compared using a multilevel mixed-effect linear regression model at the pupil-level with Core Language Standard Score as the response variable. Group allocation, baseline Core Language Standard Score and the number of children with parental consent within the school (minimisation factor, in its continuous form) will be included as fixed effects in the model.

Pupil-level fixed effects:

- Baseline Core Language Standard Score (continuous)

School-level fixed effects:

- Number of children with parental consent (continuous)

Adjustment will be made for clustering at the school level by including school as a random effect.

Model equation:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 w_i + \beta_3 I_{Ai} + u_{ij}\gamma_i + \varepsilon_{ij}$$

Y_{ij} = response of the j-th member of cluster (school) i, *i=1,…,m, j=1,…,n_i*

m = number of clusters (schools)

n_i = size of cluster (school) i

x_{ij} = Baseline Core Language Standard Score for j-th member of cluster (school) i

w_i = number of consenting pupils for cluster (school) i

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

β_0, β_1, β_2, β_3 = fixed effect parameters

u_{ij} = random effect for j-th member of cluster (school) i

γ_i = random effect parameter

ε_{ij} = residual error term for j-th member of cluster (school) i

Model assumptions will be checked as follows: the normality of the standardised residuals will be checked using a qq plot. If the model assumptions are in doubt, a sensitivity analysis will be conducted in which transformations of the outcome and/or covariate data will be tried to improve the model fit.

Secondary outcome analysis

The secondary outcomes of the individual CELF- Preschool 2 UK subtest scores and the seven CSBQ subscales will be analysed exactly as described for the primary outcome of the Core Language Standard Score, adjusting for the associated pre-test subscale score instead of the Core Language Standard Score pre-test score.

A similar approach will be taken for the analysis of the HOME total and subscale scores, except this will not adjust for the associated pre-test HOME score since this was measured post-randomisation. A sensitivity analysis adjusting for pre-test HOME score will be undertaken to assess the impact of this on the results.

The reliability of the HOME scale made up of the 6 subscale scores will be investigated using Cronbach's alpha as a measure of internal consistency.

Interim analyses

No interim analyses will be undertaken.

Subgroup analyses

Subgroup analyses for the primary outcome based on whether children are eligible for the Early Years Pupil Premium (EYPP), English as an Additional Language (EAL), and gender (separately) will be undertaken via the inclusion of the variable and an interaction term between the variable and group allocation in the primary analysis model.

Imbalance at baseline

School and pupil characteristics and measures of prior attainment will be summarised descriptively by randomised group both as randomised and as analysed in the primary analysis. School data collected at pre-test will include number of pupils moving onto primary school in 2018/19, and data from the Usual Practices survey. Pupil-level data includes gender, date of birth (to calculate age in months), ethnicity, EAL, special educational needs (SEN) and EYPP. No formal statistical comparisons will be undertaken (Senn, 1994). Continuous measures will be reported as a mean, standard deviation (SD) (and/or median, minimum and maximum) while categorical data will be reported as a count and percentage. The unadjusted difference between groups on the pre-tests for those analysed in the primary analysis will be reported as an effect size with 95% CI.

Missing data

The amount of missing baseline and outcome data will be summarised, and reasons for missing data explored and provided in the report, where available. A multilevel mixed-effect logistic regression model will be run to assess for statistically significant predictors of missing primary outcome data at the pupil-level, including all available pupil and school-level baseline data as fixed effects, and school as a random effect. Significant predictors and possible mechanisms for the missing data will be discussed in the report.

If more than 5% of randomised pupils are excluded from the primary analysis due to missing data, then the impact of missing data on the primary analysis will additionally be assessed using multiple imputation by chained equations, predicted by pre-test total Core Language Standard Score, school, allocation, number of consenting pupils, and any variables found to be significant in the 'drop-out' model described above.

A 'burn –in' of 10 will be used and 20 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.

Compliance

The EasyPeasy programme is delivered across multiple levels:
- The EasyPeasy team deliver the intervention via website application, training and text message to parents
- The Pod Leader at the school attends training and signs parents up to the programme, monitors and supports parent engagement throughout the programme through Pod management (using their web-based platform) and also potentially uses the EasyPeasy games with the nursery children in their setting
- The parents of children in the school receive the text messages and then access the games and information provided on the web-app, play the games with their child and interact with the Pod community (of other parents and the nursery staff on the web-app and face-to-face in the setting)

- The child plays the EasyPeasy games with their parent

In this trial, aspects of compliance and fidelity arise from four levels as described below:

At the top level of the technical delivery, EasyPeasy will provide a report at the end of the intervention period detailing any technical failures with the web-app or the sending of text messages to parents throughout the programme. This information will allow us to report on whether there were any technical problems or issues that may have affected the delivery of the programme as intended. The number, frequency and type of technical issues experienced will be summarised.

At the Pod Leader level, EasyPeasy will provide the evaluation team with details of which Pod Leaders attended the training. We will also collect data through the Pod Leader survey on the amount of time Pod Leaders spent supporting delivery of EasyPeasy. These data will be summarised in the final report.

At the parent level, EasyPeasy will provide information on the level of parental drop-out (number of parents who ask to be removed from the EasyPeasy programme) across the project which will be reported to give an insight into how many parents do not engage at all with the programme. Jelley *et al.* (2016) reported that 3% of parents chose to opt-out of receiving the intervention during the EasyPeasy pilot study. The number and proportion of parents who withdraw from the intervention will be reported.

It will not be possible to measure whether parents are actually playing the games with their children; however, this will be explored during the parental focus groups which will take place as part of the case studies in six settings but will not be reported as part of compliance with the programme. The summary/analysis of this data will form part of the process evaluation, and as such is not covered in this analysis plan.

A Complier Average Causal Effect (CACE) analysis for the primary outcome will be considered to account for Pod Leader attendance at the training. This variable will be measured at the nursery-level (i.e. did the nursery's Pod Leader attend training or not). An instrumental variable approach will be used, with random group allocation as the instrumental variable (Dunn, 2005). Formal CACE analysis for the primary outcome at the child-level to investigate the effects of compliance will be conducted by defining compliance of the nurseries as a dichotomous variable as 0 (Pod Leader did not attend training) and 1 (Pod Leader attended training).

Intra-cluster correlations (ICCs)

The intracluster correlation coefficient (ICC) associated with school for the post-test outcomes will be extracted from each multilevel analysis model, with the 95% CI. The ICC associated with school for the pre-test scores will also be presented with a 95% CI.

Effect size calculation

Effect sizes will be calculated based on the adjusted mean difference between the intervention and control group and the total variance (between plus within school variance), obtained from a multi-level model without covariates.

One nursery from the control group closed down post-randomisation and the 10 participants from that nursery attended one of two nurseries from the intervention group and therefore received the intervention. Sensitivity analysis will look at any impact this has on the trial outcomes.

References

Dunn, G., Maracy, M. and Tomenson, B., Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. Statistical Methods in Medical Research, 2005. 14(4): p. 369-395.

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

Howard, S.J. and Melhuish, E. 2016. An Early Years Toolbox for Assessing Early Years Executive Function Language, Self-regulation, and Social Development: Validity, Reliability, and Preliminary Norms. *Journal of Psychoeducational Assessment*, 35 (3): 255-275.

Jelley, F., Sylva, K., & Karemaker, A. 2016. EasyPeasy parenting app: Findings from an efficacy trial on parent engagement and school readiness skills. Available at: http://www.suttontrust.com/wp-content/uploads/2016/11/EasyPeasy-Evaluation_V5_FINAL.pdf

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

Slavin R, Smith D. The Relationship between Sample Sizes and Effect Sizes in Systematic Reviews in Education. Educational Evaluation and Policy Analysis. 2009 Dec;31(4):500-506. Available from, DOI: 10.3102/0162373709352369