Statistical Analysis Plan
Peep Learning Together Study
Evaluator: Queen's University Belfast
Principal investigator: Dr Sarah Miller

PROJECT TITLE	A randomised controlled trial evaluation of the Peep Learning Together Programme
DEVELOPER	Peeple, Oxford
EVALUATOR	Queen's University Belfast
PRINCIPAL INVESTIGATOR	Dr Sarah Miller
TRIAL (CHIEF) STATISTICIAN	Dr Sarah Miller
SAP AUTHOR(S)	Dr Sarah Miller, Dr Laura Dunne
TRIAL REGISTRATION NUMBER	ISRCTN64960433
EVALUATION PROTOCOL URL	*https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Round_10-_Peep_Learning_Together_Study_Protocol_amended.pdf*

SAP version history

VERSION	DATE	REASON FOR REVISION
1.0	11 May 18	N/A

Introduction

The Peep Learning Together Programme (LTP) is based on the Opportunities, Recognition, Interaction, and Modelling (ORIM) framework and aims to improve parenting skills and the quality of the home learning environment in the early years. The programme being evaluated consists of an initial home visit, and 20 one-hour sessions delivered over two terms, with groups of parents and their 3-4 year old children attending the sessions together in nursery settings. It teaches parents about how children learn and develop, to help them build on what they are already doing at home to support their child's education (home learning environment). The Programme and resources cover five topics: 1) social and emotional development; 2) communication and language; 3) early literacy; 4) early maths; and 5) health and physical development. Each session focuses on a particular topic related to children's development and includes discussion, songs and stories, as well as advice and approaches for parents to adopt at home.

It also contains comprehensive materials for practitioners which provide background theory and clear guidance for practice as well as resources for parents. Peeple provide two days of training for nursery practitioners to help them work with parents. The training covers three elements: 1) the theory on which the Programme is based, and its existing evidence-base; 2) the Programme structure and content; 3) the skills and attributes needed to deliver the Programme, and the practicalities of Programme delivery, including a practice delivery session.

Design overview

The aim of the evaluation is twofold:

- through a cluster randomised controlled trial: to determine the impact of the Peep Learning Together Programme (LTP) on the outcomes of participating parents and their three to four year-old children.
- through a process and implementation evaluation: to explore fidelity and the mechanisms through which any impact on outcomes is achieved.

Specifically, the research questions are:

Randomised controlled trial:

1. What is the impact of the Peep LTP on the communication, language and early literacy outcomes of three to four-year-old children (co-primary outcomes)?
2. What is the impact of the Peep LTP on other child related outcomes, including social and emotional development (secondary outcome)?
3. What is the impact of the Peep LTP on parent related outcomes including the home learning environment, parenting confidence and parenting stress (secondary outcomes)?
4. Is there a differential impact of the programme for children from different socio-economic backgrounds?

Process and implementation evaluation:

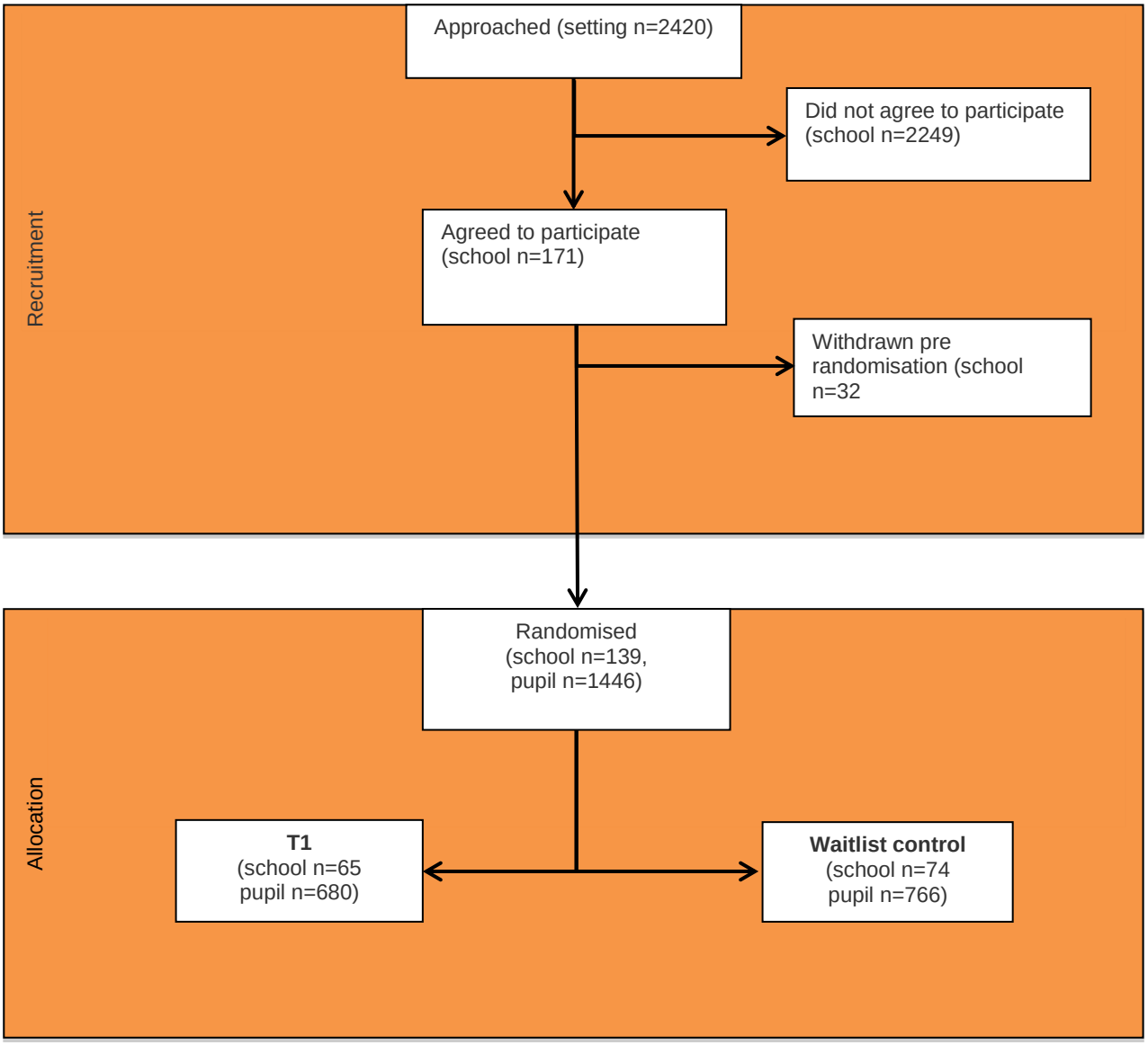
1. Was the intervention implemented with fidelity?
2. Is any variability in implementation associated with variability in outcomes?
3. Do the proposed mechanisms (e.g. the home learning environment, parental engagement) explain any link between the programme and child related primary outcomes (i.e. communication, language and early literacy skills)?
4. What were the facilitators and/or barriers to parental engagement with the programme?

Trial type and number of arms		Randomised controlled (efficacy) trial with two levels (children clustered within settings) and two arms (an intervention arm and a control arm)
Unit of randomisation		Nursery settings
Stratification variables (if applicable)		Deprivation level of the nursery setting location (as measured by the income deprivation affecting children index: IDACI)
Outcomes	primary	Language skills

	secondary	Early literacy development Social and emotional learning and communication Quality of the home learning Environment Parenting skills
Outcome sources (instruments, datasets)	primary	Clinical Evaluation of Language Fundamentals (CELF) Preschool 2UK – core language subscale
	secondary	Concepts About Print The Ages and Stages Questionnaire & the Brief Early Years Skills and Support Index The Home Learning Environment Tool to Measure Parental Self Efficacy

Follow-up

The figure below reports the number of settings recruited and the setting drop-out pre random allocation.



Sample and effect size calculations overview

Effect sizes from previous quasi experimental evaluations of the Peep LTP have varied depending on the age of the children and the outcome in question. Given this and the absence of any existing randomised controlled trials, the trial was powered to detect a minimum effect size of 0.2 of a standard deviation. The estimates reported below for the main trial (at both the protocol and randomisation stages of the study) have been calculated using Optimal Design (Version 3.01) and are based on a 2-level cluster design (where level two is the setting and level one is the child). In the table below, the parameters used to estimate the MDES for three different scenarios are described:

1. the protocol stage (prior to recruitment)
2. the allocation stage and based on actual recruited number of settings (n=139) with an average of 10 children per setting
3. for the FSM sample, assuming that 20% of the overall sample will be eligible for FSM

		PROTOCOL	RANDOMISATION	FSM
MDES		0.20	0.20	0.35
Pre-test/ post-test correlations	level 1	0.50	0.50	0.50
	level 2			
	level 3			
Intracluster correlations (ICCs)	level 1	0.10	0.10	0.10
	level 2			
	level 3			
Alpha		0.05	0.05	0.05
Power		0.80	0.80	0.80
One-sided or two-sided?		two	two	two
Average cluster size		8	10	2
Number of schools	intervention	75	65	75
	control	75	74	75
	total	150	139	150
Number of pupils	intervention	600	680	-
	control	600	766	-
	total	1200	1446	-

Between March and July 2017, Peeple recruited 171 settings to the trial, each of whom signed a memorandum of understanding agreeing to take part. Of these, 139 settings participated in baseline data collection. While this number of settings (n=139) was lower than the intended target of 150 settings, the overall power of the trial to detect a MDES of 0.2 remained unchanged because it was possible to recruit an average of ten (rather than eight) children per setting. Reasons for drop out of settings between recruitment and baseline testing (but

prior to allocation) included concern around the feasibility of implementation, which was largely linked to doubts around the ability to recruit sufficient parents (approximately ten per setting) and staffing concerns.

Upon completion of baseline data collection, 139 settings were allocated to control (n=74 settings) and intervention (n= 65 settings). Allocation was carried out using the programme Minim, minimising by setting level deprivation as measured by the IDACI decile of disadvantage. Each setting was classified as low, medium or high deprivation. Although the intervention/ control allocation is slightly imbalanced, the two groups are balanced in terms of the proportion of high, medium and low deprivation:

Control: High (36%) Med (28%) Low (35%)
Intervention: High (38%) Med (26%) Low (35%)

Until the NPD application is made, we will not know for certain how many children are eligible for FSM. Using the assumptions detailed in the table above and assuming that 20% of the sample are eligible for FSM, a MDES of 0.35 (with 80% power) is detectable with the anticipated FSM sub sample.

Analysis

Primary outcome analysis

Analysis will be conducted using Stata version 14 (Stata Corporation, College Station, Texas, USA), on an intention-to-treat basis.

The main effects of the intervention on the primary outcome - language skills as measured by the CELF core language scale - will be estimated using multilevel modelling to take account of the clustered nature of the data (where pupil is level 1 and nursery setting is level 2). Within this model, the post-test CELF score will form the dependent variable and the independent variables will include:

1. dummy variable representing whether the child was a member of the intervention or control group (coded '1' and '0' respectively)
2. pretest score for Concepts About Print
3. setting level measure of deprivation (used in the allocation process)

The main focus for the analysis will be the estimated coefficient associated with the dummy variable that represents the difference in mean scores on the respective outcome variable between the intervention and control groups, once baseline scores and other covariates are controlled for. This coefficient will be used to estimate the effect size (see below) of the programme in relation to the respective outcome variable as the standardised mean difference between the two groups at post-test (Hedges' g). The setting identifier (level 2) will be included as a random effect within the model. The equation for the model is:

$$\text{LanguageScore}_{ij} = \beta_0 + \beta_1\text{AllocationVariable}_{ij} + \beta_2\text{PreTestScore}_{ij} + \beta_3\text{Deprivation}_{ij} + u_{0j} + e_{ij}$$

Where u_{0j} is the setting level residual i.e. the effect of setting *j* on the outcome and e_{ij} is a student level residual. Residuals are assumed to be normally distributed with a mean of zero and variance σ^2_{u0}.

Secondary outcome analysis

A series of models will be estimated for each secondary outcome, which include:

Child secondary outcomes

- *Early literacy development*, measured using Concepts About Print
- *Social and emotional learning and communication*, measured using The Ages and Stages Questionnaire and the Brief Early Years Skills and Support Index

Parent secondary outcomes

- *Quality of the home learning environment*, measured using The Home Learning Environment scale
- *Parenting skills* measured using, the Tool to Measure Parental Self Efficacy

The analysis of the child secondary outcomes listed above will follow the same model specification used for the primary outcome. Thus, for each outcome, the outcome score at posttest will form the dependent variable and the independent variables will include:

1. dummy variable representing whether the child was a member of the intervention or control group (coded '1' and '0' respectively)
2. pretest score for the outcome
3. measure of deprivation (used in the allocation process)

In order to minimise the research burden on parents, pretest data for parent outcomes (quality of the home learning environment and parenting skills) were not collected. The models for these two outcomes will therefore only include two covariates:

1. dummy variable representing whether the child was a member of the intervention or control group (coded '1' and '0' respectively)
2. measure of deprivation (used in the allocation process)

Interim analyses

No interim analyses are planned.

Additional analyses

To determine whether the model is robust to the inclusion of additional covariates (known to be important in language development) further analyses will be undertaken. Specifically, for both the primary and secondary outcomes, the main models described above will be re-run, except on this occasion the following covariates will also be included: (ever)FSM, EAL and gender. The model equation becomes:

$$\text{OutcomeScore}_{ij} = \beta_0 + \beta_1\text{AllocationVariable}_{ij} + \beta_2\text{PreTestScore}_{ij} + \beta_3\text{Deprivation}_{ij} +$$

$$\beta_4\text{FSM}_{ij} + \beta_5\text{EAL}_{ij} + \beta_6\text{Gender}_{ij} + u_{0j} + e_{ij}$$

An application to the NPD will be made to obtain data relating to (ever)FSM, EAL and gender for each child.

Imbalance at baseline for analysed groups

At allocation, setting level deprivation was used in the minimisation process to ensure balance between the intervention and control groups on this variable. As such, the intervention and control groups will be compared at both baseline and analysis on this variable to examine any imbalance (using multilevel regression models to account for clustering).

In addition, core characteristics of the intervention and control groups (by setting, pupil and parent) will be compared, including:

- *Setting characteristics*: setting type, setting/school size (based on the number of children on the register), OFSTED rating;
- *Setting level pupil characteristics*: proportion of pupils eligible for FSM, attainment, proportion of pupils with English as an Additional Language (EAL), proportion of pupils with Special Educational Needs (SEN) or education and health care plans;
- *Pupil characteristics*: age, gender, EAL status, FSM eligibility;
- *Parent characteristics*: age, highest level of parental education;
- Primary and secondary outcomes measured at pre-test.

Similarly, linear and logistic multilevel regression models will be used to model the imbalance between the intervention and control groups, accounting for the clustered nature of the data. Summary statistics of these characteristics will be presented in tabular format, disaggregated by control and intervention group. Setting and pupil level attrition will also be reported from the point of randomisation to analysis. Imbalance at baseline attainment will be reported as effect size.

Missing data

Missing post-test data for language skills may occur if pupils are absent from school on the day of testing. This will be minimised by a follow-up visit to any school with pupil absences. Missing data may also occur if a pupil leaves the school completely before the post-tests are administered, if the child does not assent to participate on the day of testing, or if parents withdraw consent.

For outcomes measured using multi-item scales, there will be two types of missing data: (1) complete missing data and (2) partial missing data where some but not all items of the measure are completed. Complete missing data will be minimised by a second visit to the school to obtain data for any pupils who were absent on the main day of testing or a reminder email to parents to return the parent survey. Partial missing data will be minimised during administration by ensuring that fieldworkers double check each test prior to leaving the school on the day of data collection. The extent of missing data within each scale will be checked and cross-referenced with paper questionnaires to check for data entry errors.

For all variables, the proportion of and reason for missing data will be assessed and reported. The proportion of each outcome lost to follow-up in the control and intervention groups will be examined through cross-tabulations. If missing data are less than five percent then a complete case analysis will be undertaken in addition to multiple imputation.

Multiple imputation will be conducted as a sensitivity analysis if there are missing data. The pattern of missing data will be explored by comparing the proportion of missing data in each of the control and intervention groups in addition to exploring how missingness is related to the outcomes in question. If the data can be assumed to be MCAR then imputation is not required. If the data are assumed to be MAR, this assumption renders the missing mechanism ignorable, simplifying the imputation step whilst ensuring correct inference. The imputation model will impute data separately for the control and intervention groups and will include all relevant variables and auxiliary variables involved in the analysis and sampling design. The imputation will be performed using chained equations which fills in missing values in multiple variables iteratively by using a sequence of univariate imputation methods with fully conditional specification of prediction equations. This method accommodates arbitrary missing-value

patterns. Twenty imputations will be conducted in order to lessen the simulation (Monte Carlo) error. The analysis using the imputed datasets will then be compared to the complete case analysis. The sensitivity analysis would constitute a secondary analysis and the primary outcome analysis will be presented without the multiple imputation.

Compliance

For the purpose of this evaluation Peeple has defined compliance as 70% attendance by parents and children i.e. attending 14 of the 20 programme sessions. Attendance registers will be completed at every session by the programme facilitator, supported and mentored by Peeple. Thus it is intended that a dosage score is available as a continuous indicator for each participant, representing the number of sessions attended and ranging from a minimum of 0 (never attended) to 20 (attended every session). These data will be sent to QUB by Peeple when the programme is completed in the intervention settings. Following EEF guidance, an instrumental variable approach will be used alongside the ITT analysis to explore treatment effects in the presence of non-compliance.

Subgroup analyses

To estimate the effect of the intervention for children from deprived backgrounds the main analysis described above will be repeated on a subsample of the data identified as (ever)FSM. The models will be run for both the primary and secondary outcomes.

To determine whether the programme is more effective for children for whom English is an additional language (EAL) an interaction term will be created i.e. the product of the EAL variable, coded 1/0 and the intervention variable, coded 1/0. This interaction term will be added as an independent variable to the main analysis models for each outcome.

Effect size calculation

Hedges' g will be calculated using the following equation:

$$g = \frac{\text{coefficient}}{\text{pooled standard deviation}}$$

The standard deviation for each of the control and intervention group means will be estimated using the standard errors from the null models, estimated for each group separately for each outcome. The pooled standard deviation will be calculated using the following equation:

$$s = \frac{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}}{n_1 + n_2 - 2}$$

The standard error from the null models will also be used to calculate the 95% confidence intervals i.e. coefficient +/- (1.96 x SE).

Report tables

Minimum detectable effect size at different stages

Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	e.g. 150 (75; 75)	0.50	0.10	Minimisation on deprivation	80%	0.05	0.20
Randomisation	e.g. 139 (65; 74)	0.50	0.10	Minimisation on deprivation	80%	0.05	0.20
Analysis (i.e. available pre- and post-test)	-	-	-	Minimisation on deprivation	80%	0.05	-

Baseline comparison

Variable	Intervention group		Control group	
School-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Setting type	-	-	-	-
Ofsted rating	-	-	-	-
School-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**
Number enrolled (size)	-	-	-	-
Proportion of pupils eligible for FSM	-	-	-	-
School attainment	-	-	-	-
Proportion of pupils with EAL	-	-	-	-
Proportion of pupils with SEN	-	-	-	-
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for FSM	-	-	-	-
Gender	-	-	-	-
EAL	-	-	-	-
Pupil-level (continuous)	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**
Pretest Concepts About Print score	-	-	-	-
Pretest Ages and Stages Questionnaire Score	-	-	-	-

Primary analysis

Raw means		Effect size
Intervention group	**Control group**	

Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Language skills	-	-	-	-	-	-	-

Secondary analysis

Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Early literacy development	-	-	-	-	-	-	-
Social and emotional learning and communication	-	-	-	-	-	-	-
Quality of the home learning Environment	-	-	-	-	-	-	-
Parenting skills	-	-	-	-	-	-	-

Additional analysis, including FSM, EAL and gender as additional covariates

Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Language skills	-	-	-	-	-	-	-
Early literacy development	-	-	-	-	-	-	-
Social and emotional learning and communication	-	-	-	-	-	-	-
Quality of the home learning Environment	-	-	-	-	-	-	-
Parenting skills	-	-	-	-	-	-	-

Sensitivity analysis: complete case analysis compared and multiple imputation

Outcome	Complete case analysis Effect size			Multiple imputation		
	n in model (intervention; control)	Hedges g (95% CI)	p-value	n in model (intervention; control)	Hedges g (95% CI)	p-value
Language skills	-	-	-	-	-	-
Early literacy development	-	-	-	-	-	-
Social and emotional learning and communication	-	-	-	-	-	-
Quality of the home learning Environment	-	-	-	-	-	-
Parenting skills	-	-	-	-	-	-

Compliance analysis

	Compliers Mean language skills	**Non Compliers** Mean language skills	**All** Mean language skills
Intervention	N (%) (mean)	N (%) (mean)	N (%) (mean)
Control	N (%) (mean)	N (%) (mean)	N (%) (mean)

FSM analysis: primary outcome

Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Language skills	-	-	-	-	-	-	-

FSM analysis: secondary outcomes

Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value

Early literacy development	-	-	-	-	-	-	-
Social and emotional learning and communication	-	-	-	-	-	-	-
Quality of the home learning Environment	-	-	-	-	-	-	-
Parenting skills	-	-	-	-	-	-	-

Sub group analysis: EAL

		Raw means					
		Intervention group		Control group		Effect size	
Outcome	**Baseline attainment**	**n (missing)**	**Mean [95% CI]**	**n (missing)**	**Mean [95% CI]**	**Hedges' g [95% CI]**	**p-value of interaction term**
Primary outcome							
Language skills	EAL	-	-	-	-	-	-
	Not EAL	-	-	-	-	-	
Secondary outcome							
Early literacy develop-ment	EAL	-	-	-	-	-	-
	Not EAL	-	-	-	-	-	
SEL and commu nication	EAL	-	-	-	-	-	-
	Not EAL	-	-	-	-	-	
	EAL	-	-	-	-	-	-

Quality of the HLE	Not EAL	-	-	-	-	-	
Parenting skills	EAL	-	-	-	-	-	-
	Not EAL	-	-	-	-	-	