

Statistical Analysis Plan

[PACT]

Evaluator (institution): Durham University

Principal investigator(s): Victoria Menzies, Dr Helen Cramman



Education
Endowment
Foundation

Template last updated: March 2018

PROJECT TITLE	Parents and Children Together (PACT)
DEVELOPER (INSTITUTION)	Manchester University
EVALUATOR (INSTITUTION)	Durham University
PRINCIPAL INVESTIGATOR(S)	Victoria Menzies, Dr Helen Cramman
SAP AUTHOR(S)	Professor Adetayo Kasim, Dr Nasima Akhter, Victoria Menzies, Dr Helen Cramman, Professor Jochen Einbeck, Jessica Hugill, Paivi Eerola, Dr Lyn Robinson-Smith
Trial Statisticians	Dr Nasima Akhter, Professor Adetayo Kasim, Professor Jochen Einbeck
TRIAL DESIGN	Two-arm randomised controlled trial with random allocation at pupil level
Trial Type	Efficacy Trial
PUPIL AGE RANGE AND KEY STAGE	3-4 year olds (preschool)
NUMBER OF SCHOOLS	48
NUMBER OF PUPILS	480
PRIMARY OUTCOME	Improvements to language development (assessed using a language latent variable from the LanguageScreen assessment) at delayed post-test
SECONDARY OUTCOME	Improvements to different components of language development (assessed by LanguageScreen subtest scores) at delayed post-test Improvements to the child's home learning environment measured by the Home Learning Environment Index at post-test School readiness, measured by BESSI at delayed post-test

SAP version history

VERSION	DATE	REASON FOR REVISION
1.2 [<i>latest</i>]		
2.1	09/11/2021	Adaptation in analysis plan due to Covid-19 impacting intervention delivery and completion of face-to-face assessments with children.
1.0 [<i>original</i>]	03/03/2020	

Table of contents

SAP version history	2
Table of contents	3
Study rationale and background	3
Research questions	5
Study Design	5
Randomisation	6
Sample size calculations	7
Outcome measures	8
PRIMARY	8
SECONDARY	10
Statistical Analysis	12
Imbalance at baseline	12
PRIMARY OUTCOME	12
SECONDARY OUTCOME	13
SUBGROUPS ANALYSIS	13
ADDITIONAL ANALYSIS	13
Longitudinal follow-up analyses	13
Missing data	13
CACE ANALYSIS.....	14
Intra-cluster correlations (ICCs).....	15
Effect size calculation	15
SOFTWARE	15
REPORTING OF RESULTS	15
Cost evaluation	16
Data protection	16
References	16

Study rationale and background

Vocabulary acquisition is a key element of early infant development and continues to be an important factor throughout childhood. Bergelson and Swingley (2012) reported that babies appear to start learning the sound forms of whole words within the first few months of life and they understand the meanings of several common nouns from the age of six months. At around the age of 18 months, young children's vocabulary begins to expand rapidly and it is estimated that they learn words at a rate of one every two waking hours; a trend that will continue to adolescence (Pinker, 1994). In addition to vocabulary acquisition, infants need to learn about the features of spoken language such as where words begin and end, and realise that these units carry a meaning. This phonological knowledge underpins vocabulary acquisition and growth.

Moving on to learning to read, Harrison (2004) suggested that children need different types of knowledge as precursors:

- Knowledge and understanding of the world; knowledge of how our language works;
- Knowledge of conventions of print; phonological awareness; decoding, oral reading fluency and reading comprehension are beginning to be acquired by many children by 5 years of age.

Evidence indicates that parenting and educational environment in the early years have a powerful influence on language development. The quality of the home learning environment and educational resources within the home are important factors (Melhuish *et al.* 2008b) and there is a link between this quality and socio-economic status (Foster *et al.* 2005). We observe children from disadvantaged backgrounds entering school with lower levels of attainment than their more socioeconomically advantaged peers (Tymms *et al.* 2014) and this trend persists throughout primary school (Merrell, Little and Coe, 2014); development and skills at the start of school are predictive of later outcomes (See, for example, Tymms, Merrell and Bailey, 2017).

Parents and Children Together (PACT) provides teaching sessions and materials for parents/carers to use with their children to develop their language skills. Previous research suggests that through its structured programme to enhance the home learning environment, PACT could positively impact on the quality of the home learning environment, leading to gains in language development. While various dimensions of language skills are generally measured using standard tests, research has also shown that language skills reflect a unitary construct and can be measured using a latent variable sharing common variances of these dimensions (NELI Evaluation - Dimova *et al.*, 2020; MacDonald, 2013). Latent variables have been used in assessing effectiveness of language interventions in different contexts, including assessment of parental interventions (Burgoyne, 2018). Selection of test components to construct a latent language skill variable is generally theory driven, but studies have used different combinations of these tests (Dimova, *et al.*, 2020; Burgoyne, 2018; West *et al.* 2021). Recent research assessed correlation between two unitary constructs of latent language variables and observed high correlations between them (West *et al.*, 2021).

The PACT trial was originally planned as a multi-site trial to be implemented in the usual school setting with children from a school randomised to intervention and control groups and data collected using the same tools during pre-test and post-test stage, allowing standard estimation of effect of the intervention. At the pre-test stage, several researcher led assessments (The Clinical Evaluation of Language Fundamentals (CELF) Preschool 2; British Picture Vocabulary Scale – 3 (BPVS-3); Action Picture (APS) test; Listening comprehension), measuring dimensions of language skills were used. These tools will be used for calculating a language latent variable. However, during COVID-19 the usual school setting was hugely disrupted, affecting the delivery and assessment of the intervention. The COVID-19 restrictions in schools and in universities made it impossible to have the researcher-delivered face-to-face assessment of children in schools, as was specified in the original project protocol (**Table 3**). To deal with this unanticipated issues, in consultation with EEF, an app-based language skill assessment tool 'LanguageScreen' was used at the delayed post-test period, which was suitable for delivery by staff in schools. COVID-19 restrictions meant that assessment was only possible 12-months after the intervention (delayed post-test) and not at the immediate post-test period. The calculation of a latent

language variable is still possible using the LanguageScreen app-based assessment (West et al. 2021), replacing the original post-test language latent variable. COVID-19 also had some implication on assessment of home learning environment. Further details are given in relevant sections.

Research questions

1. What is the impact of the PACT intervention on the language development of participating children, as measured using a language latent variable, assessed by LanguageScreen app delivered by school staff at the end of the first year of school (reception year) (12 months after the intervention period)? [Primary Outcome]
2. What is the impact of PACT on the Expressive vocabulary, Receptive vocabulary, Listening comprehension, and Sentence repetition (LanguageScreen sub-set items) of participating children? [Secondary Outcome]
3. What is the impact of PACT on the home learning environment of participating children at the end of nursery and after two months in school as measured using the Home Learning Environment Index? [Secondary Outcome]
4. What is the impact of the PACT intervention on the school readiness of participating children as measured at the end of the first year of school? [Secondary Outcome]

Research questions 1, 2, and 4 will be investigated when the children are at the end of their first year in school – 12-months after the scheduled end of the PACT programme (delayed post-test). Research question 3 (using parent completed measures) will be investigated immediately after the scheduled end of the PACT Programme and 5 months later.

Study Design

The study design is a two-armed randomised controlled efficacy trial delivered under ideal conditions with allocation at pupil level. Pupils will be allocated equally to either the intervention (pupils allocated to receive the PACT programme) or control group (pupils allocated to 'business as usual' plus equivalent incentive cost of materials (approximately £130) in books to parents/carers on completion of the immediate post-test). The proposed within school randomisation will be more powerful than cluster randomisation if there is negligible heterogeneity in intervention effects between school and if there is no dilution of the intervention effects as a result of contamination between intervention and control groups. *The books included in the programme could be passed between parents/carers in the PACT group to the control group, however the developers advised the core of the programme is the accompanying activities and resources, which after initial completion would not be particularly useful as some of the materials are single use. Additionally, the staff training could encourage change of practice within the setting, however the developers intend the training to focus on the theory of the programme and how best to support parents/carers in its delivery, none of which is expected to create new*

knowledge significant enough to influence classroom based practice. The summary of trial design is presented in Table 1. Unlike most EEF trials, the primary outcome is a latent construct of language development measured by a combination of LanguageScreen app sub-scale scores (Expressive vocabulary, Receptive vocabulary, Listening comprehension, and Sentence repetition).

Table 1: Study design summary

Trial type and number of arms	Two-armed randomised controlled efficacy trial
Unit of randomisation	Pupil level
Stratification variables (if applicable)	Pre-test completeness
Primary outcome	Variable measure (instrument, scale)
	Language development A latent language variable combining LanguageScreen sub-scale scores (Expressive vocabulary, Receptive vocabulary, Listening comprehension, and Sentence repetition).
Secondary outcome(s)	variable(s) measure(s) (instrument, scale)
	(1) Early Language skills, (2) Home Learning Environment, (3) School Readiness (1) LanguageScreen sub-scale scores : Expressive vocabulary (EV), Receptive vocabulary (RV), Listening comprehension (LC), and Sentence repetition (SR) (2) Home Learning Environment Index, (3) BESSI

Randomisation

The trial statistician who is not involved in the recruitment of schools or parents/carers completed randomisation independently. All participating pupils were allocated into one of the two groups (intervention or control) on a 1:1 ratio.

Ideally to minimise any potential bias in completing pre-testing and provide maximum data for the study all pre-testing would have been completed before randomisation. However, due to the tight timeframe for fitting in the 30-week intervention during the school year, and difficulties with children being absent from nurseries when researchers visited, randomisation needed to take place before all pre-testing was complete. Children were included in the study if they had completed a minimum of the CELF expressive vocabulary and the CELF sentence structure assessments during one assessment session with a researcher. Where a researcher had sat and attempted to complete the assessments with the child but

the child was not compliant in completing the pre-testing assessments these children were also still included in the randomisation. To account for any potential systematic bias which may have explained whether pre-test was complete or not at date of randomisation, randomisation was undertaken to ensure the two groups (control and intervention) are balanced on pre-test completion status as follows:

- (1) 'Pre-test complete' which will include all children who complete all baseline assessments, which form the originally planned latent language variable (CELF Expressive Vocabulary, CELF Sentence Structure, Listening Comprehension, BPVS-3, APT) and the CELF Word Structure measure.
- (2) 'Partial pre-test complete' which includes at a minimum both the completion of CELF expressive vocabulary and the CELF sentence structure
- (3) No pre-test data available

Whilst randomising based on pre-test scores would be desirable, this was not possible due to timing of the pre-test assessments, the required data entry time and delivery of the PACT programme requiring the majority of the academic year (30 weeks).

Sample size calculations

Based on the maximum capacity for delivery by the developer 48 schools were recruited and randomised assuming 480 children in total (i.e., 10 per school), along with 5% Type I error, 80% Power, 10% intra-school correlation, 60% pre-post test correlation and 2-sided test. Using these assumptions, the sample size will detect a minimum difference of 0.18 standard deviations between the PACT and control group. The intra-cluster correlation of 10% was based on the average value observed in EEF trials (Xiao *et al*, 2016). We have assumed 60% correlation between pre-test and post-test score as Burgoyne *et al.* (2018a, 2018b) found previously. The sample size was estimated using Optimal Design software for person randomised multisite trials that will allow detecting a minimum effect size (REF).

The implication of COVID-19 for this trial meant that our sample size calculation assumed the same pre-test and post-test assessments and a shorter time between these assessments as per the original protocol and original design of the study. While we have some evidence (West *et al.*, 2021) to assume that there will be a good correlation between the pre-test latent variable using the researcher delivered in-depth assessments and the revised delayed post-test latent variable using the shorter teacher-delivered LanguageScreen app-based assessment, it is not known for certain that this will be the case. Furthermore, the 12-month delay in the collection of primary outcome data has led to a greater level of attrition, as pupils moved from nurseries to different schools. Altogether, this may result in a reduced power than originally expected. A power calculation will be included in the final report.

Table 2: Summary of sample size calculations for the actual number of schools recruited (as of June 2019)

	OVERALL
MDES	0.18

Pre-test/ correlations	post-test	level 1 (pupil)	0.60
		level 2 (class)	
		level 3 (school)	
Intracluster (ICCs)	correlations	level 2 (class)	
		level 3 (school)	10%
Alpha			0.05
Power			0.8
One-sided or two-sided?			Two-sided
Average cluster size			10
Number of schools	Intervention		48
	Control		48
	Total		48
Number of pupils	Intervention		240
	Control		240
	Total		480

Outcome measures

PRIMARY

The proposed primary outcome measure is the language latent variable at delayed post-test which will use the four subtests' scores from the LanguageScreen app: Expressive vocabulary (EV), Receptive vocabulary (RV), Listening comprehension (LC), and Sentence repetition (SR). This assessment is administered to the child by a member of staff in the child's school using the LanguageScreen app on a tablet. Full instructions are included within the app for the delivery of the assessment with the need for training. Verbal instructions and questions for the child are played aloud through the app rather than given by the administrator. This should minimise variability in the delivery of the assessment across participants. The subtests are always presented in the same order. The assessment involves:

- a) Receptive Vocabulary – the child is asked to select which of four pictures matches a spoken word on a series of items by tapping the correct picture on the screen and the app automatically marks whether this was correct (23 items).
- b) Expressive Vocabulary – the child is asked to name a series of pictures. The test administrator judges and records in the app whether the child answers correctly (24 items).
- c) Listening Comprehension – the child listens to three stories which are each followed by a series of questions about the story. The test administrator uses app guidance to judge whether the answer is correct and records this in the app (16 items).
- d) Sentence Repetition – the child is asked to repeat verbatim a series of spoken sentences. The test administrator judges and records in the app whether child did this (14 items).

The four assessments are presented in a set order and take around 25 minutes to complete altogether. Data from the app is uploaded to the LanguageScreen website automatically and results are generated automatically by LanguageScreen. A standardised and raw score for each subtest is provided as well as an overall raw and standardised score.

In the COVID-19 context, this standardised app-based assessment was suggested by EEF and is also used by the Nuffield Early Language Intervention (NELI) intervention trial. The original pre-test and immediate post-test language latent variable were due to use the same test items of a) Expressive Vocabulary and Sentence Structure subtests of Child Evaluation of Language Fundamentals (CELF) Preschool IIUK ; b) British Picture Vocabulary Scale (BPVS), d) *The Renfrew Action Picture Test* (APT; Renfrew, 2003; information and grammar scores) and d) Listening Comprehension, as was used by Burgoyne et al. (2008).

The new delayed post-test LanguageScreen latent variable will replace the original immediate post-test language latent variable. The pre-test latent language variable will still be developed (as planned) for inclusion in the model to adjust for pre-test scores.

Given that there were unexpected COVID-19 disruption, we consider this is the best possible analytical approach that will still be reasonable for pre-test adjustment. Our justification is that, firstly, a recent study observed a very high correlation ($r = 0.95$) between two latent constructs of language skills: one used LanguageScreen sub-tests and another utilised scores from four items: the CELF2 EV and Recalling sentences subtests, and APT information and grammar test scores (West et al., 2021). West uses the same post-test latent variable as in the current study. However, the pre-test latent variables are different, with some similarities. The similarities exist between West (2021) and this study for three out of the four items in the pre-test latent variables, which are 1) LanguageScreen EV similar to CELF2 EV; 2) LanguageScreen RV similar to BPVS RV); and 3) LanguageScreen LC similar to Listening Comprehension measure.

The latent variable will be constructed using the structural equation modelling (SEM). Model fit will be assessed using the following criteria: root mean square error of approximation (RMSEA) <0.08 ; standardized root mean square residual (sRMR) <0.08 ; comparative fit index (CFI) ≥ 0.90 ; and Tucker-Lewis index (TLI) ≥ 0.95 (Hu and Bentler, 1999; Kline, 1998). If the planned model is not a good fit, modification necessary for a parsimonious model will be made in consultation with EEF. Both pre-test and delayed post-test latent variable construct will apply this approach.

As defined by the developer's previous PACT trial (Burgoyne 2018a, 2018b), the pre-test latent variable will allow for comparison to the pre-test results of this trial with another trial and then allow it to be utilised in the primary outcome analysis. This will be measured with:

- (a) The **Clinical Evaluation of Language Fundamentals Preschool 2 UK (CELF-Preschool 2 UK)** includes subscale scores of sentence structure and expressive vocabulary. CELF-Preschool 2 UK provides a measure for expressive and receptive language skills in young

children. This is a standardised and validated assessment with the proposed age group and UK sample.

- (b) The **British Picture Vocabulary Scale – 3 (BPVS-3)**, is a standardised measure of receptive vocabulary appropriate to 3 year olds. The programme activities specifically target vocabulary learning and involve increased exposure to a variety of books and resources. This measure consists of a set of pictures from which the child is asked to point to the picture representing a given word.
- (c) The **Action Picture Test (APT)** is a standardised test that requires children to give samples of spoken language in response to picture stimuli. The test considers grammatical structures used and the expressive vocabulary used. The test is suitable to use with children between the ages of 3 and 8 and provides normed scores. It is quick and simple to administer and inexpensive to purchase.
- (d) **Listening Comprehension** as measured by a task developed by the developer team used previously across a variety of projects. In this assessment children listen to a recording of a short story. The tester then asks eight comprehension questions and records the child's response verbatim for later scoring by the research team using detailed scoring guidance. The test takes approximately three minutes to administer. Materials for this will be provided by the developer at no additional cost.

SECONDARY

The LanguageScreen individual **subscale** scores (EV, RV, LC, SR), measured at delayed post-test will form a secondary outcome. These subtests measure the child's vocabulary, listening and sentence recalling skills. Apart from that, Brief Early Skills and Support Index (BESSI) and Home Learning Index (HLE) will form additional secondary outcomes.

The **Brief Early Skills & Support Index** (or **BESSI**) questionnaire (measured at the delayed post-test) will be used to evaluate school readiness. BESSI is a simple but reliable questionnaire which assesses how well children are making the transition to school. This questionnaire has been developed and validated for reception and nursery children. Teachers will be asked to complete this questionnaire for children in the project at delayed post-test. The total score for this measure will be used in the analysis.

The **Home Learning Environment Index** (HLE; Melhuish et al, 2008) which was developed as part of the EPPE study, and has been used in several large studies including the Millennium Cohort Study, National Evaluation of Sure Start (NESS) and a study of the Home Learning Environment by the Scottish Government (Melhuish, 2010). The HLE asks parents/carers to report the frequency of seven routine activities which are conceptually linked to learning (including being read to, going to the library, playing with numbers, painting and drawing, being taught letters, being taught numbers and songs/poems/rhymes). These seven items were positively linked with predicting under and over achievement at aged 5 (Melhuish *et al.* 2008). Frequency of the seven activities is coded on a 0 to 7 scale). Previously, this index was used in surveys conducted over the phone, however, for the PACT trial, the questions were added to the usual practice survey for parents/carers at the beginning of the trial and immediately following the teaching period (i.e. during the immediate post-testing period). The total score (out of 49) for the seven activities will be the score used for the analysis. This measure was

adapted for delivery at immediate post-testing due to the Covid-19 pandemic and the restrictions which may have changes how parents interacted with their children in the home. Therefore, parents were asked to give answers for two timepoints – reflecting back to February 2020 before the pandemic and answering for their current experience in June 2020 when they were completing the survey. In addition, parents were asked to complete the survey for a final time in November 2021. The HLE data available for June 2020 & Nov 2021 will be used in analysis.

Table 3 List of assessments at each assessment time-point and revisions from original protocol.

Pre-test		Immediate Post-test		Delayed Post-test	
Original protocol	Revised post-covid Protocol	Original protocol	Revised post-covid protocol	Original protocol	Revised post-covid protocol
Language Measures with pupils in schools					
CELF sentence structure (MU - RA delivered)		CELF sentence structure (MU - RA delivered)	Not possible due to COVID-19 context	CELF sentence structure (MU - RA delivered)	LanguageScreen (DU – school staff delivered)
CELF expressive vocabulary (MU – RA delivered)		CELF expressive vocabulary (MU – RA delivered)	Not possible due to COVID-19 context	CELF expressive vocabulary (MU – RA delivered)	
Listening comprehension (MU – RA delivered)		Listening comprehension (MU – RA delivered)	Not possible due to COVID-19 context	Listening comprehension (MU – RA delivered)	
BPVS III (MU – RA delivered)		BPVS III (MU – RA delivered)	Not possible due to COVID-19 context	BPVS III (MU – RA delivered)	
Renfrew Action Picture Test (MU – RA delivered)		Renfrew Action Picture Test (MU – RA delivered)	Not possible due to COVID-19 context	Renfrew Action Picture Test (MU – RA delivered)	
CELF word structure (MU – RA delivered)		CELF word structure (MU – RA delivered)	Not possible due to COVID-19 context	YARC Letter Sound Knowledge, Early Word Reading, Sound Deletion (MU – RA delivered)	
Parent completed measures					
Home Learning Environment Index (HLE)	Home Learning Environment Index (HLE)	Home Learning Environment Index (HLE)	Home Learning Environment Index (HLE) in June 2020 & Nov 2021	-	
Reception staff completed measures					

				Brief Early Skills & Support Index (BESSI)	Brief Early Skills & Support Index (BESSI)
--	--	--	--	--	--

Statistical Analysis

The primary outcome and secondary outcomes will be analysed using the principles of intention to treat, meaning that all schools and pupils will be analysed in the group they were randomised to, irrespective of whether or not they actually get the PACT programme. Statistical significance will be assessed at the 5% level. Results will be reported according to the EEF reporting template.

Imbalance at baseline

All the baseline data will be presented by intervention and control group using descriptive statistics. Cross-tabulation of background characteristics (including gender and by Early Years Pupil Premium (EYPP)) will be presented. We will also perform cross-tabulation between the pre-test complete status (Pre-test completed, Partial pre-test completed and No pre-test data available) and the intervention status. Additional data on pupils and school characteristics will be described. For continuous variables, we report means and standard deviations, and for, for categorical data, counts and percentages. Note that no effect size will be presented for baseline data.

PRIMARY OUTCOME

The primary outcome considering a latent language variable derived by combining 4 variables from scores on LanguageScreen sub-tests' (1.Expressive vocabulary, 2.Receptive vocabulary, 3.Listening comprehension, 4. Sentence repetition), will be analysed on a continuous scale using a confirmatory factor analysis model. The pre-test language latent variable (described above) will be used for baseline adjustment using ANCOVA model. The effect size and its confidence intervals will be estimated as standardised factor loadings from confirmatory factor analysis models. A similar modelling approach was used in the previous trial (Burgoyne 2018). This modelling approach enables estimation of impacts of PACT across the different components of language development as measured by the latent outcome. It assumes that the language skills may be better assessed as a latent construct that uses shared variance of the sub-tests and can reflect important elements of language skills that may be difficult to measure relying on observed variables. Sensitivity analysis will be performed using multilevel structural equation modelling to test whether the estimated effects of the intervention are constant across schools. Missing data in the pre-test and outcomes measures will be accounted for using full information maximum likelihood estimation techniques (Cham *et al.*, 2017).

SECONDARY OUTCOME

All non-latent variable secondary outcomes will be analysed using multilevel models with school and school-by-intervention as random effects. The effect size and the associated confidence intervals will be calculated using conditional variance of the outcome data to ensure consistency of results with the latent variable model, where the confidence interval for the effect of the intervention will be based on conditional variance. The immediate and delayed impacts of the PACT intervention on the secondary outcomes will be analysed using a multilevel model accounting for intra-school correlation. Exploratory analysis comparing the difference between the immediate impact of the intervention and five-month follow up for the HLE will also be reported.

SUBGROUPS ANALYSIS

All the outcome data will be analysed by Early Years Pupil Premium (EYPP) eligibility model. Alongside fitting the latent variable model separately for subgroup category of those who received EYPP, interaction model will also be considered. Effect size for pupils eligible for EYPP will be reported in accordance with EEF requirement.

ADDITIONAL ANALYSIS

We will analyse the individual data included in the latent variable model using multilevel models to investigate whether some outcomes are more sensitive to the intervention than the other outcomes.

Longitudinal follow-up analyses¹

No longitudinal data collection has been planned. The analysis dataset (excluding LanguageScreen results) will be archived and longitudinal analysis could be conducted in the future by linking the data with NPD. LanguageScreen delayed post-test data will be pseudonymised when archived and will not be available to link with NPD as there is no agreement obtained with the LanguageScreen developer to link the LanguageScreen dataset with administrative data.

Missing data

Missing data will be assessed both at pre-test at post-test. Pre-test missing data will be presented using cross-tabulation between missing data completeness status (completed, partially completed and no pre-test data) and the intervention groups. We would also investigate percentage of missing data in each of the individual components of the primary outcome data, and further analysis regarding imputations will be done when >5% of the outcome data are missing. The latent variable approach with the full information maximum likelihood estimation implicitly assumed that the underlying mechanism for the missing data does not depend only on the observed data. This missingness mechanism is commonly termed 'missing at random' and the full information maximum likelihood estimation estimate

the parameters of the latent variables conditioning on the observed data for each of the latent outcomes. It also assumed that all outcomes are linearly related with each other and are multivariate normally distributed, which enables it to condition missing data on observed data assuming multivariate normal distribution (Charm *et al.*, 2017). In order to check whether the assumption of missing at random holds, we would also perform multiple imputation on the composite outcomes and then apply latent variable model to estimate the impact of the intervention. We would expect that results from multiple imputation and the full information maximum likelihood estimation lead to similar conclusion if the underlying missingness mechanism is missing at random. We would consider ten imputations for each outcome using chained equations or the Markov chain Monte Carlo (MCMC) method, which allows non-monotone imputation between pre-test and post-test data (Jakobsen *et al.*, 2017). To impute pre-test data for a particular outcome, the pre-test scores for other outcomes will be used in the imputation model. However, both pre-test and post-test data will be used in the imputation model for any of the post-test outcomes. The imputation approach will be sequential such that all the pre-test scores will first be imputed and then they would be used in turn to impute the post-test outcomes. Note that the effect size from each of the imputation will be presented as range of values for sensitivity analysis. We would not consider a dropout model for the multiple imputation because of the nature of the latent variable model. The collective missing data is more important in the latent variable model rather than individual dropout model. Lastly, we would use all the available data on the latent primary outcome for missing data imputation.

CACE ANALYSIS

Self-reported compliance data will be collected from parents/carers in both the intervention and control groups as part of the standard intervention delivery and through the usual practice survey. The self-reported compliance data will measure to what extent has each child in the intervention group adhered to the required sessions of the intervention. Compliance data on number of sessions delivered based on data submitted by parents about each session using the PACT app and paper record forms (total number of sessions completed) will be used in a Complier Average Causal Effect (CACE) analysis. The CACE analysis will be implemented using instrumental variable approach by comparing the outcomes between the intervention group and control group with a focus on random variation in compliance data. In other word, it will assess conditioning on number of PACT session/home support, what is the impact of PACT intervention on language development (Pokropek, 2016). The model will assume that the impact of compliance on language development is only through the intervention and usual practice (i.e. intervention groups). This means that there would be no impact of the intervention if a child had no session of PACT. Under this assumption, instrumental variable would be incorporated as an additional node in a structure equation model. However, the instrumental variable will be considered observed instead of latent with the disadvantage that it does not

allow for testing whether the self-reported compliance data is a valid instrument (Pokropek, 2016).

Intra-cluster correlations (ICCs)

There is no explicit estimation of ICCs in a latent variable model. However, we will estimate ICCs for the analysis of the individual outcome data using multilevel models. The pre-test estimation of ICCs will be based on a model with only the overall mean and with schools as random effects. The estimation of ICCs for post intervention data will be done with and without fixed effects, but with schools as random effects. ICCs will be computed at school level.

Effect size calculation

The effect size for the primary outcome will be obtained directly from confirmatory factor analysis model as standardized factor loading on intervention variable. This estimated effect size will be compared with effect size for the individual measurable primary outcome data using Hedges' g effect size from a multilevel model defined as

$$ES = \frac{\widehat{\mu}_T - \widehat{\mu}_C}{\sqrt{\sigma_w^2 + \sigma_s^2 + \sigma_I^2}}$$

Where $\widehat{\mu}_T - \widehat{\mu}_C$ is the adjusted average difference between the intervention and control groups. σ_w^2 is residual variance, σ_s^2 denotes between school variance and σ_I^2 denotes the variance of school by intervention effects. The effect size will be computed using both conditional and unconditional variance.

SOFTWARE

We will analyse the data using statistical software R (most updated version at the time of analysis) and SAS 9.4.

REPORTING OF RESULTS

All results will be reported using the EEF template as shown in Table 4.

Table 3: Template for reporting results

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

Cost evaluation

Data on intervention costs will be collected from the developers as well as from schools participating in the PACT Lead interviews, through interviews as part of the process evaluation, and will be used to conduct a cost evaluation in line with recent guidance from the EEF.

Data protection

The legal basis for processing the personal data accessed and generated by the trial is Public Task covered by GDPR Article 6 (1) (e) public task, which states that; “the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law.” No special category data will be collected as part of this project.

References

Bergelson, E. and Swingley, D. (2012) At 6–9 months, human infants know the meanings of many common nouns, *PNAS*, 109 (9), 3253 – 3258. www.pnas.org/cgi/doi/10.1073/pnas.1113380109 Article retrieved on 06/06/12

Boereboom, J. & Cramman, H. (2018). Primary school entry assessment in New Zealand. *NZ International Research in Early Childhood Education Journal*, 21(1), 47 - 61. Retrieved from <https://www.childforum.com/research/2018-nz-international-early-childhood-education-journal/1564-primary-school-entry-assessment-in-new-zealand.html>

Burgoyne, K., Gardner, R., Whiteley, H., Snowling, M.J. & Hulme, C. 2018a. Parents and Children Together (PACT): Evaluating a parent-delivered oral language enrichment programme for pre-school children. Final Report. The Nuffield Foundation.

Burgoyne, K., Gardner, R., Whiteley, H., Snowling, M.J. & Hulme, C. 2018b. Evaluation of parent-delivered early language enrichment programme: evidence from a randomised controlled trial. *The Journal of Child Psychology and Psychiatry*, 59, 5, pp.545-555.

Cham, H., Reshetnyak, E., Rosenfeld, R. and Breitbart, W. (2017) Full Information Maximum Likelihood Estimation for Latent Variable Interactions with Incomplete Indicators. *Multivariate Behav Res.*, 52(1): 12–30. doi:10.1080/00273171.2016.1245600.

Copping, L.T., Cramman, H., Gott, S., Gray H., & Tymms, P., (2016). Name writing ability not length of name is predictive of future academic attainment. *Educational Research*, Vol. 58, Iss. 3, 2016. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/00131881.2016.1184948>

Cramman, H., Gott, S., Little, J., Merrell, C., Tymms P., & Copping, L. T., (2018): Number identification: a unique developmental pathway in mathematics? *Research Papers in Education*. Published online: 1st November 2018. Retrieved from <https://doi.org/10.1080/02671522.2018.1536890>

Dimova, S., Ilie, S., Rosa Brown, E., Broeks, M., Culora, A., & Sutherland, A. (2020). The Nuffield Early Language Intervention Evaluation report. Education Endowment Foundation. Available at https://educationendowmentfoundation.org.uk/public/files/Nuffield_Early_Language_Intervention.pdf

West, G., Snowling, M. J., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., McLachlan, H., & Hulme, C. (2021). Early language screening and intervention can be delivered successfully at scale: evidence from a cluster randomized controlled trial. *Journal of child psychology and psychiatry, and allied disciplines*, 10.1111/jcpp.13415. Advance online publication. <https://doi.org/10.1111/jcpp.13415>

Jakobsen, J.C., Gluud, C., Wetterslev, J and Winkel, P. (2017) When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology*, 17:162. DOI 10.1186/s12874-017-0442-1

Harrison C. (2004) *Understanding Reading Development*, Sage Publications, London.

Hu, L. t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.

Hulme, C., Nash, H.M., Gooch, D., Lervag, A. & Snowling, M.J. (2015) The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science*, 26, 12, pp.1877-1866.

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

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

Kline, R. B. 1998. *Principle and practice of structural equation modeling*, New York, Guilford.

MacDonald, M. (2013). How language production shapes language form and comprehension, *Frontiers in Psychology*, 4, 226 <https://doi.org/10.3389/fpsyg.2013.00226>

- Melhuish, E.C., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., & Phan, M. 2008a. 'Effects of the Home Learning Environment and preschool center experience upon literacy and numeracy development in early primary school'. *Journal of Social Issues*, 64, 157-188.
- Melhuish, E.C., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., Phan, M., & Malin, A. 2008b. 'Preschool influences on mathematics achievement'. *Science*, 321, 1161-1162
- Melhuish, E.C. (2010). *Impact of the Home Learning Environment on Child Cognitive Development: Secondary Analysis of Data from 'Growing Up in Scotland'*. Report available on the Scottish Government Website. <http://www.gov.scot/Publications/2010/04/27112324/0>
- Merrell, C., & Tymms, P. 2011. 'Changes in Children's Cognitive Development at the Start of School in England 2001 – 2008'. *Oxford Review of Education*, 37 (3).
- Merrell, C., & Bailey, K. 2012. 'Predicting achievement in the Early Years: How important is personal, social and emotional development?' *On-line Educational Research Journal*, June 2012.
- Merrell, C., Little, J. & Coe, R. (2014) *Is the Attainment Gap among Primary Aged Children Decreasing?* In *Harnessing what works in eliminating educational disadvantage: A tale of two classrooms*, Eds. Wood, C. and Scott, R. Pub. Demos: London.
- Merrell, C., and Tymms, P. 2016. 'Assessing Young Children: Problems and Solutions. In *Understanding What Works in Oral Reading Assessments*'. Montreal: UNESCO Institute for Statistics (UIS). DOI: <http://dx.doi.org/10.15220/978-92-9189-196-2-en>
- Pinker, S. 1994a. *The language instinct*. New York: Morrow.
- Pokropek, A. (2016) Introduction to instrumental variables and their application to Large-scale assessment data. *Large-scale Assess Educ*, 4:4. DOI 10.1186/s40536-016-0018-2
- Tymms, P., Merrell, C., Hawker, D., & Nicholson, F. 2014. 'Performance indicators in primary schools: A comparison of performance on entry to school and the progress made in the first year in England and four other jurisdictions'. Research Report for the Department for Education.
- Tymms, P., Merrell, C. & Bailey, K. 2017. *The Long Term Impact of Effective Teaching. School Effectiveness and School Improvement*. Published online, 14th December 2017 <http://dx.doi.org/10.1080/09243453.2017.1404478>
- West, G., Snowling, M. J., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., ... & Hulme, C. (2021). Early language screening and intervention can be delivered successfully at scale: evidence from a cluster randomized controlled trial. *Journal of Child Psychology and Psychiatry*.