Statistical Analysis Plan for FFT Reciprocal Reading

Queen's University Belfast

INTERVENTION	FFT Reciprocal Reading
DEVELOPER	FFT Literacy
EVALUATOR	CESI
TRIAL REGISTRATION NUMBER	ISRCTN81582662
TRIAL STATISTICIAN	Dr Liam O'Hare, Dr Patrick Stark, Professor Paul Connolly
TRIAL CHIEF INVESTIGATOR	Dr Liam O'Hare
SAP AUTHOR	Dr Liam O'Hare & Dr Patrick Stark
SAP VERSION	1
SAP VERSION DATE	08/02/2018

Protocol and SAP changes

Original Statements from EEF protocol and updated information

"The primary outcomes of the trial will be two attainment indicators from the New Group Reading Test (NGRT 1. Overall reading score and 2. Comprehension subscale score). However, this decision will be reviewed on submission of the statistical analysis plan (SAP) to EEF and before the analysis takes place."

We have included these as primary outcomes in the SAP below.

"Exact measures for secondary outcomes are to be decided, but these should align with the comprehension behaviours, awareness and culture outlined in the logic model (the decoding sub-scale from the NGRT will also be a secondary outcome). Therefore self- regulation, thinking strategies and attitudes towards reading have been suggested (these secondary outcomes will also be considered for assessment at the teacher level in the MOU and post-test survey instruments). These will be further discussed in the development phase and will need to be added to a protocol amendment in summer 2017."

These have now been decided and are included in the SAP below.

Table of contents

Introduction

This protocol is for a pragmatic randomised controlled trial of the Reciprocal Reading training provided by the Fischer Family Trust Literacy (FFT). The Reciprocal Reading training (RR) is a workforce development programme that supports practicing Teachers/Teaching Assistants develop and deliver comprehension instruction in mainstream UK Key Stage 2 settings. Two versions of the programme are being evaluated in this study: 1. A whole class version for UK year 4 pupils; and 2. A targeted version for UK year 5 and year 6 pupils.

Teachers involved in the universal version of the programme deliver guided reading sessions of 20-30 minutes in length, once a week, for a minimum of 12 weeks over two academic terms to Year 4 students. Over the course of the intervention students encounter many texts. The strategy is flexibly used and adapted to the texts. In the universal intervention the reading situation is whole-class, as opposed to small-group in the original version. Whole-class or whole-group reciprocal reading sessions are followed by individual tasks based on the reading sessions in the form of book journal activities.
Teachers and Teaching Assistants involved in the targeted version of the RR programme deliver guided-reading sessions of 20-30 minutes in length to small groups of Year 5 and Year 6 students who have been identified as having reading comprehension skills which are relatively weaker than their decoding skills. The frequency in the targeted version is twice a week for at least 12 weeks over two academic terms.

The protocol outlines a research design that will assess whether the RR programme over a school year can improve a number of specific outcomes in intervention group pupils, in a sample of schools experiencing higher than average levels of disadvantage. The primary aim of the programme is to improve reading comprehension ability. The primary outcomes being investigated in this trial are overall literacy and reading comprehension as measured by the comprehension New Group Reading Test [NGRT] subscale). The NGRT test was chosen as it provides standardised and distinct measures of both reading comprehension and reading accuracy. This allows analysis of both overall literacy and the main intended area of progress from the programme: reading accuracy. It is important to analyse overall literacy as a primary outcome as it is a potentially high impact pupil outcome. Reading comprehension is also being analysed as a primary outcome, due to specificity matching of outcome to programme, as this is the main area of development targeted by the intervention Secondary programme outcome effects will also be assessed on reading accuracy (NGRT accuracy subscale) and comprehension pre-cursors at the child (metacognition & self-regulation), staff (comprehension instruction behaviour and awareness), and school level (comprehension ethos). The study will also include a process evaluation using qualitative data (interviews, observations and focus groups of staff and pupils), quantitative implementation data (engagement, fidelity, delivery quality and dosage) and demographic variables (Gender, FSM, SEN).

In summary the study will address the following research questions:

1. What is the impact of the FFT RR training programme at post-test on reading outcomes (primary and secondary) in pupils participating in a universal/whole-class version?

2. What is the impact of the FFT RR training programme at post-test on reading outcomes (primary and secondary) in pupils participating in a targeted version which involves pupils identified with having relatively good reading accuracy with relatively weaker reading comprehension skills?

3. What evidence is there to support the pathways for change in both primary and secondary outcomes as proposed in the logic model for both the universal and targeted versions of the programme?

4. What is the relationship between outcomes and implementation factors for both the universal and targeted versions of the programme?

5. What does the implementation process data tell us about how the programme was implemented?

Study design

Description of population including eligibility criteria

Recruitment for the study took place from March 2017 to June 2017 and 114 schools were recruited by QUB (and signed a Memorandum of Understanding). The universal intervention is being delivered to pupils in Year 4. The targeted intervention is being delivered to pupils in Years 5 and 6. For the universal programme all children in the respective year are being given the intervention by a RR trained instructor. The eligibility criteria for the targeted version are children who are poor at reading comprehension but relatively good at decoding. The children in this category were selected by teachers in June 2017 using guidance and materials provided by FFT literacy. Teachers were provided with a document that allowed them to compare each child in their class to two sets of criteria. The first set of criteria describes the reading skills possessed by a child who could be classified as having good decoding skills. The second set of criteria describes difficulties that may be experienced by a child who struggles with reading comprehension. By comparing each child with these two sets of criteria, teachers identified children who are good at decoding, but struggle with reading comprehension. Children who matched the provided criteria for having good decoding skills

and poor reading comprehension skills will be deemed to be eligible for the targeted intervention. Teachers selected 6 pupils from each Year 5 and 6 class for the targeted intervention.

Description of trial design

1. Cluster RCT Evaluation: The main outcomes of the RR training will be evaluated using an intention to treat cluster RCT with control group analysis. The RCT will test for changes in pupils primary reading outcomes (reading comprehension, overall reading) as well as a range of secondary outcomes. Both the universal whole class version and the targeted small-group version of the implementation will be tested. Any changes in the intervention group receiving the FFT RR training will be measured against the control group who do not receive the treatment. Pre-test NGRT data was collected in July 2017. Post-test NGRT data will be collected in July 2018.

2. Process evaluation. A process evaluation will supplement the RCT to measure implementation factors. It will seek to assess dosage, reach, fidelity and quality. To help assess this all teachers and teaching assistants delivering the programme will complete a questionnaire, an audit tool will be administered by FFT and observations and interviews will be carried out in ten case study schools.

Randomisation

Randomisation was carried out in July 2017 after pre-test in June 2017.

Stratification was used as part of the randomisation process. Stratification improves the precision of the estimates by helping to ensure that the treatment indicator is orthogonal to the other covariates (Cox and Reid 2000).

In this case we conducted minimisation through the QMinim software package. Minimisation is a well-recognised approach that uses algorithms to ensure a balance on certain covariates between the control and intervention schools at baseline, and is especially useful when randomising a small number of cases (Torgerson and Torgerson, 2007). Minimisation was used to ensure the schools were as evenly matched as possible. A number of school level covariates was used in the matching process specifically: reading comprehension score (NGRT passage comprehension), % FSMEver, reading accuracy score (NGRT sentence completion). Median values were calculated for each of these characteristics to determine a mid-cut point and the creation of dichotomous variables, coding schools as “High” or “Low” for each of these characteristics. These variables were then entered into QMinim for each school; and all variables given a weight of one with the exception of reading comprehension which was double weighted as an important predictor of the outcome of interest. This randomisation process resulted in 51 schools being assigned to the Intervention condition and 49 being assigned to the Control condition.

Calculation of sample size

Sample size calculation

Sample size calculations were carried out prior to school recruitment. The RCT will assess both the whole group approach and the more targeted version of the programme and, as such, needs to be adequately powered for both. As the targeted approach, with fewer pupils, requires the higher number of schools, the following power calculations have been based on the targeted approach.

Effect sizes for literacy interventions evaluated through a good quality RCT design would tend typically be in the range of 0.2-0.3 (e.g. Biggart et al, 2013; Borman et al, 2007; Tymms et al, 2011) and the more conservative effect size of 0.2 has been used in the power calculation/s below.

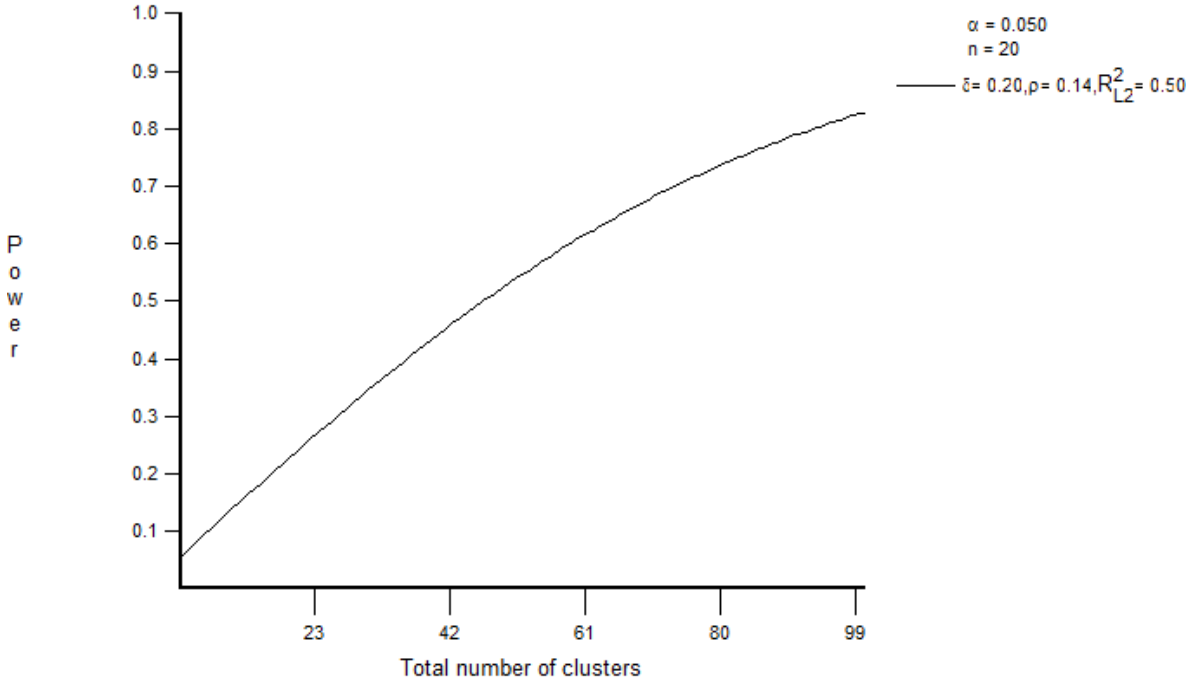


Figure 1. shows a power calculation for the RR trial provided by Optimal Design software using estimates of ES=0.2 ρ=0.05; ICC = .14 (FSM quintile 3 KS2 reading score from EEF's ICC guidance[1]); r2=0.50 (due to having a pre-test of NGRT); average class size n=7. Study power is shown according to number of clusters.

This calculation suggests a total sample size of 94 schools (clusters) to detect a significant effect if present with a power of .8.

[1] https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/ICC_2015.pdf

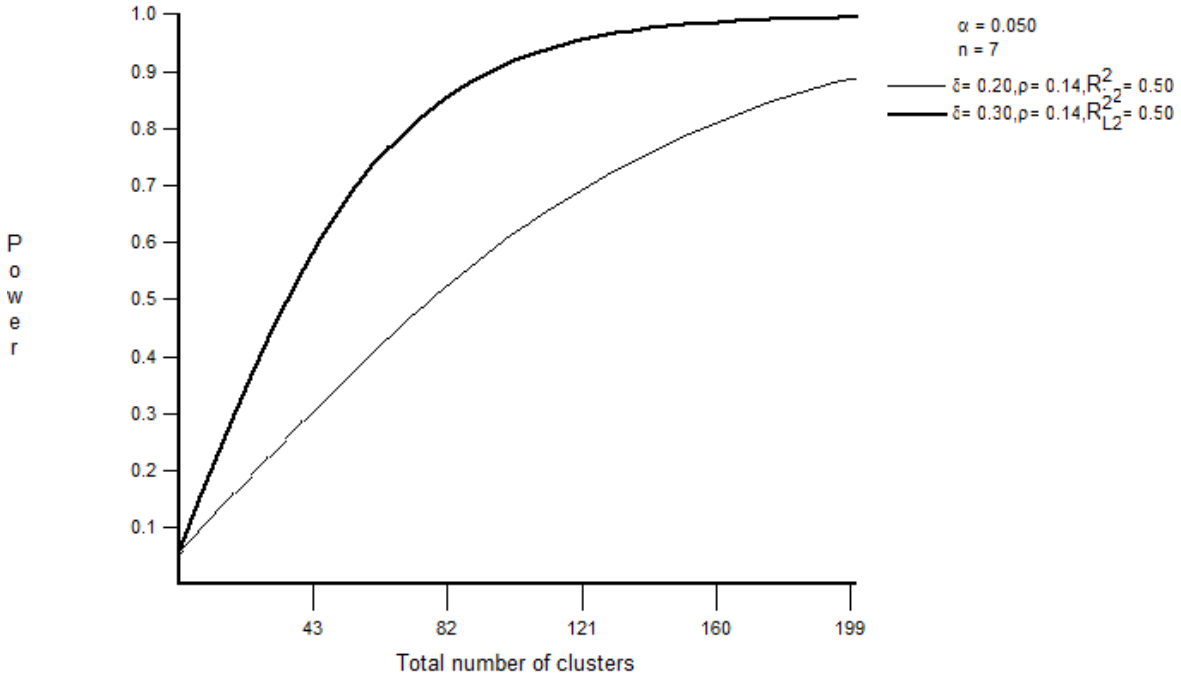


Figure 2. Study power according to number of clusters (for everFSM students)

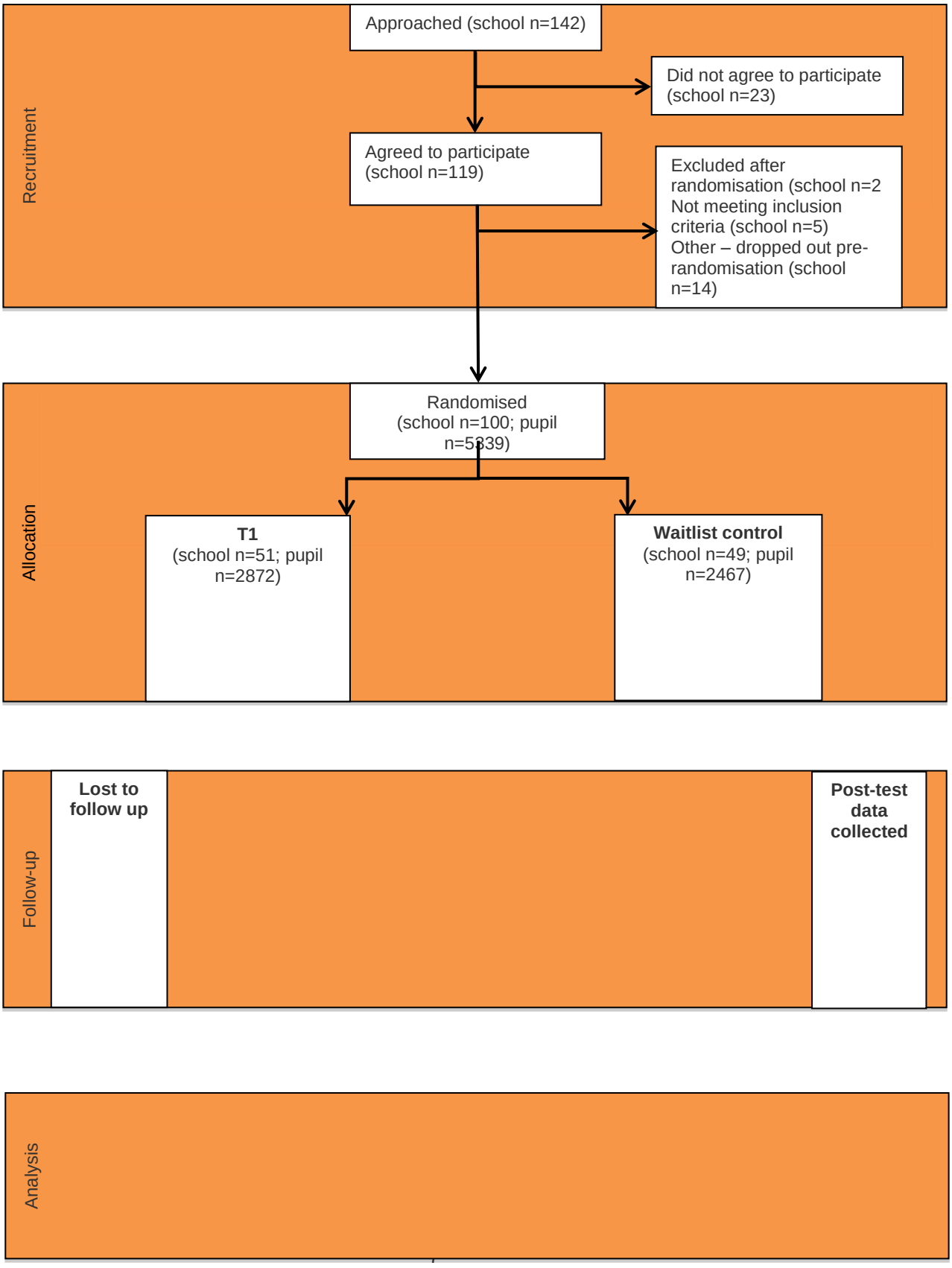
This calculation suggests a total sample size of 152 schools (clusters) to detect a significant effect of ES= 0.2 if present with a power of .8. Or a total sample size of 72 schools (clusters) to detect a significant effect of ES= 0.3 if present with a power of .8.

In conclusion, the nature of the trial with 100 schools (which is a limiting factor due to project team capacity to deliver training) permits the well powered assessment of outcomes of the programme in comparison to control schools if an effect of 0.2 is present and a subgroup analysis of everFSM pupils if an effect of 0.3 is present.

Updated MDES have been included based on 100 schools recruited in the report tables.

Follow-up

Figure 3 shows the numbers of schools participating in the trial from the point of recruitment, pre-testing and randomisation through to November 2017.



Outcome measures

PRIMARY OUTCOME

The primary outcomes of the trial are two attainment indicators from the New Group Reading Test (NGRT 1. Overall reading score and 2. Comprehension subscale score). The NGRT is an adaptive test which has high reliability and measures both reading accuracy and comprehension. Overall reading is being analysed as a primary outcome as it may be seen by teachers as an accessible and easily interpretable outcome of an intervention. Reading comprehension is being analysed as a primary outcome as it is the specified target of the programme and the most valid way to evaluate programme success. At pre-test the NGRT was delivered (digitally) by schools under exam conditions. At post-test the NGRT will be administered (digitally) by the evaluation team under exam conditions. The NGRT provides two tests A & B, and a different test will be used at pre-test and post-test to avoid practice effects. The digital NGRT form A was delivered at pre-test. Digital NGRT form B will be delivered at post-test. Overall reading scale score will be used as the score for Overall reading ability. Paragraph Comprehension scale score will be used as the score for reading comprehension. These scores provide a valid measure of each of the primary outcomes. There are two primary outcomes for this trial and there is much debate in the literature over cut-offs and criteria for correcting for multiple comparisons. There are similarly harmful aspects to correcting when unnecessary as to not correcting when necessary (Wason, Stecher & Mander (2014). The analysis in this study meets the criteria to not require correction for multiple testing due a small number of planned comparisons (Armstrong, 2014).

Table 1: Primary Outcome Measures

Primary Outcome	Measure	Level of Measurement	Number of items	Reliability (Cronbach's Alpha)
Reading Comprehension	New Group Reading Test – Passage Comprehension	Pupil	27	0.9
Overall reading ability	New Group Reading Test – Total Score	Pupil	47	0.9

SECONDARY OUTCOMES

The secondary outcomes of the trial are Reading accuracy score from the Sentence Completion subtest of the NGRT, pupil meta-cognition and pupil comprehension behaviour. The pupil meta-cognition and pupil comprehension behaviour measures were developed by the evaluation team. These bespoke measures were developed to provide specificity matching to the intervention as no previously used standardised measures would be appropriate for this. Reliability will be calculated for the pupil meta-cognition and pupil comprehension behaviour measures at post-test as these are bespoke measures. This will include removal of unreliable items (i.e., items that reduce scale reliability) in the calculation of the scale score. If the scale does not meet the conventional threshold for reliability (i.e., alpha = .7) this will be highlighted in the report and readers will be encouraged to interpret findings with caution. These measures have been designed to closely reflect the principles of the Reciprocal Reading intervention. The school level comprehension ethos questionnaire will be completed by head teachers and is designed to measure the prioritisation and focus of the teaching of reading comprehension in the school,

e.g. if there is an ethos within the school that reading comprehension is an aspect of development to be strongly emphasised within teaching.

No pilot phase for secondary measures was in the protocol for this RCT. We understand that EEF requires previously standardised measures for *primary* outcomes, but also that the need for specificity matching of secondary outcomes and implementation factors requires the need to use bespoke measures where appropriate and has been the norm with numerous previous trials, e.g. development of bespoke pupil/teacher engagement measures for programmes. Also, secondary outcome measures are used to explore theory of change. Therefore it is important to match programme outcomes even more closely than primary outcomes. Overall, the programme is judged on primary outcomes and theory of change explained by secondary outcomes

Table 2: Secondary Outcome Measures

Secondary Outcome	Measure	Level of Measurement	Number of Items	Reliability (Cronbach's Alpha)
Reading Accuracy	New Group Reading Test – Sentence Completion	Pupil		0.9
Pupil Meta-Cognition	Post-test Pupil Meta-Cognition Survey	Pupil		Calculated at post-test.
Pupil Comprehension Behaviour	Post-test Pupil Comprehension Behaviour Survey	Pupil		Calculated at post-test.

The additional analysis will explore the role of implementation factors in primary and secondary outcome change. Regression analyses will be conducted using each implementation factor as a predictor of primary and secondary outcomes. These implementation factors and sources of measurement are detailed below in Table 3.

Table 3: Implementation Factor Measures and National Pupil Database data for Additional Analysis

Implementation Factor	Measure	Level of Measurement
Teacher Comprehension awareness, behaviour and attitude	Bespoke measure teacher comprehension pedagogy development questionnaire (RR version)	Teacher
School comprehension ethos	MOU & Principal survey	Head teacher

Dosage	24 week implementation record	Teacher
Pupil engagement	Pupil survey at post-test	Pupil
Teacher engagement	Teacher Survey at post-test	Teacher
SEN	National Pupil Database	Pupil
FSM	National Pupil Database	Pupil
Gender	National Pupil Database	Pupil
KS1 literacy	National Pupil Database	Pupil

Analysis

PRIMARY INTENTION-TO-TREAT (ITT) ANALYSIS

Analysis will be carried out using STATA. Analysis will be conducted on an intention-to-treat basis. The main effects of the intervention will be estimated using multilevel random-effects linear regression modelling to take account of the clustered nature of the data and a series of models will be estimated for each outcome (where pupil is level 1 and school is level 2). Multilevel random effects models allow for estimates of variance of both levels.

Firstly, analysis will be conducted for the universal intervention with the NGRT comprehension score, overall reading score forming the dependent variable and the independent variables including a dummy variable representing whether the child was a member of the intervention or control group (coded '1' and '0' respectively) and pupils' baseline scores at pre-test.

Secondly, this analysis will be repeated for the targeted intervention.

See Tables 4 and 5 below for summaries of regression analyses for primary analyses. The syntax for these models is included in Appendix 1.

Table 4 - Summary table of regression models for primary analysis of universal intervention

Model	*Dependent variable*	*Independent variable 1*	*Independent variable 2*
1	Post-test – NGRT Overall reading score	Group – Control or Intervention	Pre-test – NGRT Overall reading score
2	Post-test – NGRT Reading Comprehension	Group – Control or Intervention	Pre-test – NGRT Reading Comprehension

Table 5 - Summary table of regression models for primary analysis of targeted intervention

Model	*Dependent variable*	*Independent variable 1*	*Independent variable 2*
3	Post-test – NGRT Overall reading score	Group – Control or Intervention	Pre-test – NGRT Overall reading score
4	Post-test – NGRT Reading Comprehension	Group – Control or Intervention	Pre-test – NGRT Reading Comprehension

IMBALANCE AT BASELINE FOR ANALYSED GROUPS

The analysis used to help determine whether attrition has led to imbalance at baseline will be multi-level regression models. These will test for differences in reading comprehension, FSMEver % and reading accuracy between the Control and Intervention groups. Imbalance at baseline will be reported as effect size. Effect size will be calculated using the equation described in "Effect Size Calculation" on Page 15 of this SAP.

MISSING DATA

Schools who drop out of delivering the intervention will be encouraged to allow post-testing and will still be included in an intention-to-treat analysis. If the proportion of missing data is higher than 5%, a 'missing at random' data analysis will tell us whether imputation is required. If so, data will be imputed using multiple imputation which will be presented as a sensitivity analysis. The missing at random analysis will be carried out in Stata and will determine if the pattern of missing data is related to the primary outcome. Schools that were pre-tested but then dropped out of the trial will be included as ITT.

NON-COMPLIANCE WITH INTERVENTION

Dosage will be calculated using the naturally occurring data from FFT as recorded by teachers in the weekly lesson frequency recording tool given to teachers by FFT on their initial training day. This survey asks teachers to record the number of lessons and the average length of lessons on a weekly basis. The progress in completion of this recording tool will be checked during the schools follow-up training visit by FFT. QUB fieldworkers will collect this survey from teachers on the day of the post-test.

Dosage will be included as an independent variable in the exploratory analysis of programme implementation factors. We will also explore the use of CACE analysis[2]. Year groups of participants will be coded as compliant/non-compliant based on comparing their teachers' reported dosage with the minimum required dosage. FFT have reported that minimum dosage should be two sessions of twenty minutes per week for twelve weeks for the Targeted Intervention, and one session of twenty minutes per week for twelve weeks for the whole class intervention. Year 4 (whole class intervention) groups with a reported dosage of under 240 minutes, and year 5 and 6 groups (targeted intervention) with a reported dosage of under 480 minutes will be coded as non-compliant. A regression analysis will then be conducted on the same scores as the primary analysis, but also including compliance level as a predictor.

SECONDARY OUTCOME ANALYSES

Secondary analyses will investigate differences between the control and intervention groups in post-test reading accuracy, pupil meta-cognition and pupil comprehension behaviour, controlling for pre-test reading score. This analysis will be carried out for the universal intervention and for the targeted intervention.

Table 6 - Summary table of regression models for secondary analysis of universal intervention

Model	Dependent variable	Independent variable 1	Independent variable 2
5	Post-test – NGRT Reading Accuracy	Group – Control or Intervention	Pre-test – NGRT Reading Accuracy
6	Pupil meta-cognition	Group – Control or Intervention	Pre-test NGRT reading comprehension
7	Pupil comprehension behaviour	Group – Control or Intervention	Pre-test NGRT reading comprehension

[2] https://ac.els-cdn.com/S0022440516000133/1-s2.0-S0022440516000133-main.pdf?_tid=f1d8784e-d366-11e7-9754-00000aab0f26&acdnat=1511782655_cd42d10bb6b3ca2034cd318cc31334cd

Table 7 - Summary table of regression models for secondary analysis of targeted intervention

Model	Dependent variable	Independent variable 1	Independent variable 2
8	Post-test – NGRT Reading Accuracy	Group – Control or Intervention	Pre-test – NGRT Reading Accuracy
9	Pupil meta-cognition	Group – Control or Intervention	Pre-test NGRT reading comprehension
10	Pupil comprehension behaviour	Group – Control or Intervention	Pre-test NGRT reading comprehension

ADDITIONAL ANALYSES

Additional analysis will be conducted to explore the pathways for change in both primary and secondary outcomes as proposed in the logic model (see Protocol for logic model). This will be carried out for both the universal intervention and the targeted intervention. The original protocol had not fully specified implementation factors. These have been developed and updated since the original protocol and will be included in the updated protocol.

Table 8 - Summary table of regression models for exploratory analysis of the pathways in the logic model for the universal intervention

Model	Dependent variable	Independent variable 1	Independent variable 2	Independent variable 3
11	Pupil meta-cognition and comprehension behaviour	Group – Control or Intervention	Teacher Comprehension awareness, behaviour and attitude.	School Comprehension ethos
12	Post-test NGRT reading comprehension	Group – Control or Intervention	Pupil meta-cognition	Pupil comprehension behaviour
13	Post-test Overall NGRT reading ability	Group – Control or Intervention	Pre-test NGRT reading comprehension	N/A
14	Post-test Overall NGRT reading ability	Group – Control or Intervention	Pre-test NGRT reading accuracy	N/A

Table 9 - Summary table of regression models for exploratory analysis of the pathways in the logic model for the targeted intervention

Model	Dependent variable	Independent variable 1	Independent variable 2	Independent variable 3
15	Pupil meta-cognition and comprehension behaviour	Group – Control or Intervention	Teacher Comprehension awareness, behaviour and attitude.	School Comprehension ethos
16	Post-test NGRT reading comprehension	Group – Control or Intervention	Pupil meta-cognition	Pupil comprehension behaviour
17	Post-test Overall NGRT reading ability	Group – Control or Intervention	Pre-test NGRT reading comprehension	N/A

18	Post-test Overall NGRT reading ability	Group – Control or Intervention	Pre-test NGRT reading accuracy	N/A

The second part of the additional analysis is the analysis of implementation factors and outcome change. These models will be carried out for the intervention group only, to explore programme implementation of both the universal intervention and the targeted intervention.

Table 10 - Summary table of regression models for exploratory analysis of programme implementation factors for universal intervention

Model	*Dependent variable*	*Independent variable 1*	*Independent variable 2*	*Independent variable 3*
19	Pupil meta-cognition and comprehension behaviour	Pupil engagement	Teacher engagement	Dosage
20	Post-test NGRT reading comprehension	Pupil engagement	Teacher engagement	Dosage
21	Post-test overall NGRT reading ability	Pupil engagement	Teacher engagement	Dosage

Table 11 - Summary table of regression models for exploratory analysis of programme implementation factors for targeted intervention

Model	*Dependent variable*	*Independent variable 1*	*Independent variable 2*	*Independent variable 3*
22	Pupil meta-cognition and comprehension behaviour	Pupil engagement	Teacher engagement	Dosage
23	Post-test NGRT reading comprehension	Pupil engagement	Teacher engagement	Dosage
24	Post-test overall NGRT reading ability	Pupil engagement	Teacher engagement	Dosage

SUBGROUP ANALYSES

A subgroup analysis will repeat the primary outcome analysis for only pupils in receipt of FSM. We have since updated the protocol to include analysis of only FSM as a subgroup and this will be included in the amended protocol to be published in the near future.

Table 12 – Summary table of regression models for subgroup analysis of Primary Outcomes for FSM pupils for Whole Class Intervention

Model	*Dependent variable*	*Independent variable 1*	*Independent variable 2*
25	Post-test – NGRT Overall reading score	Group – Control or Intervention	Pre-test – NGRT Overall reading score

26	Post-test – NGRT Reading Comprehension	Group – Control or Intervention	Pre-test – NGRT Reading Comprehension

Table 13 – Summary table of regression models for subgroup analysis of Primary Outcomes for FSM pupils for Targeted Intervention

Model	*Dependent variable*	*Independent variable 1*	*Independent variable 2*
27	Post-test – NGRT Overall reading score	Group – Control or Intervention	Pre-test – NGRT Overall reading score
28	Post-test – NGRT Reading Comprehension	Group – Control or Intervention	Pre-test – NGRT Reading Comprehension

Finally, a correlational analysis of relationship between KS1 data from the National Pupil Database and NGRT overall literacy will be conducted.

EFFECT SIZE CALCULATION

The standard error will be used to calculate 95% confidence intervals for the coefficient, using the formula:

$$95\%\ \text{CI} = \text{coefficient} +/- (1.96 \text{ x standard error})$$

$$\text{Upper limit of CI} = \text{coefficient} + (1.96 \text{ x standard error})$$

$$\text{Lower limit of CI} = \text{coefficient} + (1.96 \text{ x standard error})$$

Effect size will be calculated for the primary outcomes – overall reading score and reading comprehension score. Effect size (Hedges' g) will be calculated as the standardised mean difference between the control and intervention groups, using the pooled standard deviation. The pooled standard deviation will be calculated using the formula:

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

Hedges' g will then be calculated as:

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

Report tables

Baseline Comparison

Variable	Intervention group n (missing)	Intervention group Mean (SD)	Control group n (missing)	Control group Mean (SD)	Hedges g [95% CI]	p
Pre-test NGRT overall reading						
Pre-test NGRT reading comprehension						
Pre-test NGRT reading accuracy						
FSM%						

Minimum detectable effect size (MDES) at different stages of trial for whole school intervention

Stage	N (schools)	Correlation between pre-test and post-test	ICC	Blocking or pair matching	Power	Alpha	MDES
Protocol	94	.7	.14	Pair matching	.8	.05	.2
Pre-test[3]	100	.85	.14	Pair matching	.8	.05	.15

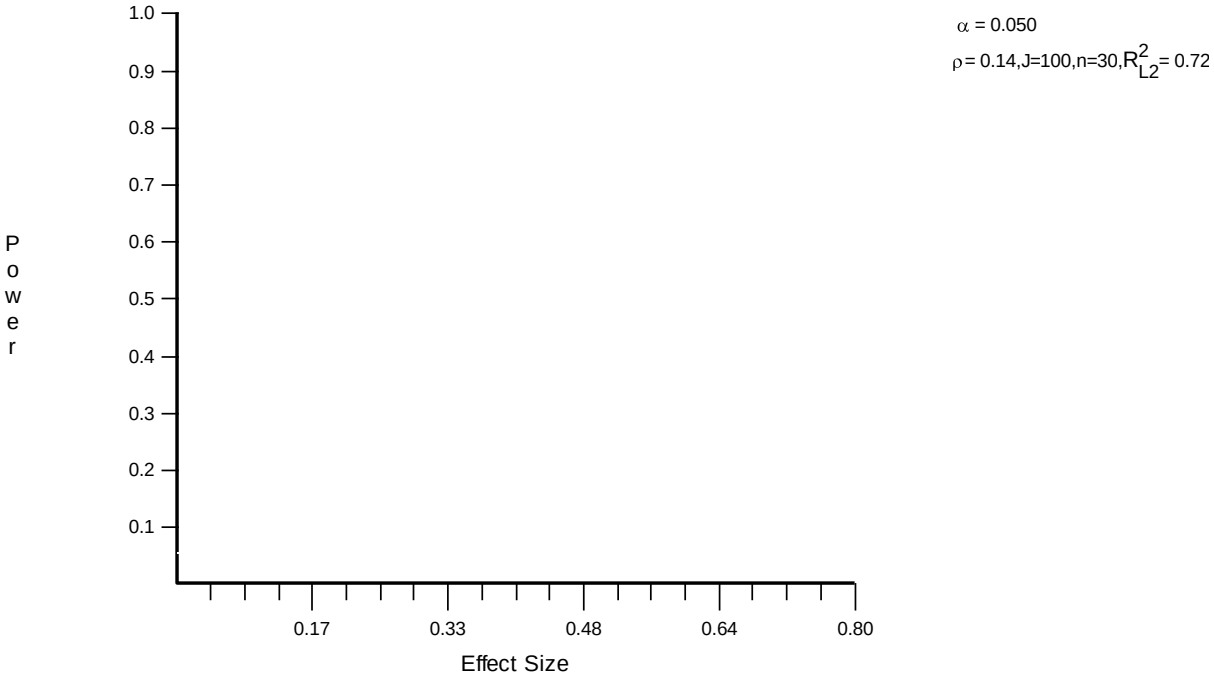


Figure 4 is a plot of power and effect size with alpha = .05, pre-test post-test correlation = .85, ICC = .14, number of clusters = 100, n per cluster = 30.

Minimum detectable effect size (MDES) at different stages of trial for targeted intervention

[3] NGRT pre-test and post-test correlation was updated after recruitment stage to reflect the recruited sample size and the reported test-retest correlation for NGRT (Granada Learning, 2013). N per cluster was included in whole class calculation as 30 for one class, and 6 for the targeted intervention as teachers were asked to select 6 targeted pupils.

Stage	N (schools)	Correlation between pre-test and post-test	ICC	Blocking or pair matching	Power	Alpha	MDES
Protocol	72	.7	.14	Pair matching	.8	.05	.3
Pre-test	100	.85	.14	Pair matching	.8	.05	.24

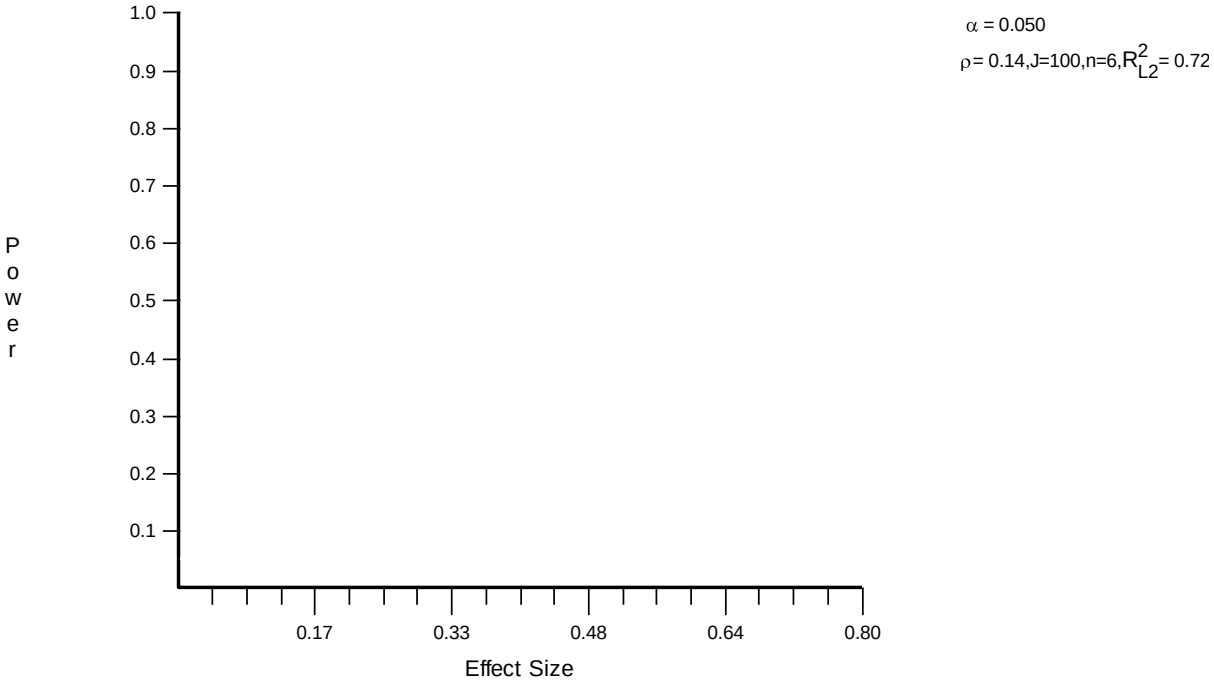


Figure 5 is a plot of power and effect size with alpha = .05, pre-test post-test correlation = .85, ICC = .14, number of clusters = 100, n per cluster = 6.

Primary Analysis

Results Table 1 - Primary analysis for Whole Class Intervention – Models 1 & 2

Outcome Primary outcome	Covariate	Raw means Intervention group		Raw means Control group		N in model (intervention; control)	Effect size Hedges g [95% CI]	p
		n (missing)	Mean [95% CI])	n (missing)	Mean [95% CI]			
NGRT Overall reading score	Pre-test NGRT Overall reading score							
NGRT Reading Comprehension	Pre-test NGRT Reading Comprehension							

Results Table 2 - Primary analysis for Targeted Intervention – Models 3 & 4

Outcome Primary outcome	Covariate	Raw means Intervention group		Control group		N in model (intervention; control)	Effect size Hedges g [95% CI]	p
		n (missing)	Mean [95% CI])	n (missing)	Mean [95% CI]			
NGRT Overall reading score	Pre-test NGRT Overall reading score							
NGRT Reading Comprehension	Pre-test NGRT Reading Comprehension							

Secondary Analyses

Results Table 3 – Secondary analysis for Whole Class Intervention - Model 5

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NRGT reading accuracy						
Group						
Pre-test NGRT reading accuracy						

Results Table 4 – Secondary analysis for Whole Class Intervention – Model 6

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Pupil meta-cognition						
Group						
Pre-test NGRT reading comprehension						

Results Table 5 – Secondary analysis for Whole Class Intervention – Model 7

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Pupil comprehension behaviour						
Group						
Pre-test NGRT reading comprehension						

Results Table 6 – Secondary analysis for Targeted Intervention - Model 8

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NRGT reading accuracy						
Group						
Pre-test NGRT reading accuracy						

Results Table 7 – Secondary analysis for Targeted Intervention – Model 9

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Pupil meta-cognition						
Group						
Pre-test NGRT reading comprehension						

Results Table 8 – Secondary analysis for Targeted Intervention – Model 10

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Pupil comprehension behaviour						
Group						
Pre-test NGRT reading comprehension						

Additional Analyses

Exploratory Analyses

Results Table 9 - Exploratory analysis for Whole Class Intervention – Model 11

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Pupil meta-cognition and comprehension behaviour						
Group						
Teacher Comprehension awareness, behaviour and attitude						
School Comprehension ethos						

Results Table 10 - Exploratory analysis for Whole Class Intervention – Model 12

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT reading comprehension						
Group						
Pupil meta-cognition						
Pupil comprehension behaviour						

Results Table 11 - Exploratory analysis for Whole Class Intervention – Model 13

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT Overall reading						
Group						
Pre-test NGRT reading comprehension						

Results Table 12 - Exploratory analysis for Whole Class Intervention – Model 14

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT Overall reading						
Group						
Pre-test NGRT reading accuracy						

Results Table 13 - Exploratory analysis for Targeted Intervention – Model 15

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Pupil meta-cognition and comprehension behaviour						
Group						
Teacher Comprehension awareness, behaviour and attitude						
School Comprehension ethos						

Results Table 14 - Exploratory analysis for Targeted Intervention – Model 16

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT reading comprehension						
Group						
Pupil meta-cognition						
Pupil comprehension behaviour						

Results Table 15 - Exploratory analysis for Targeted Intervention – Model 17

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT Overall reading						
Group						
Pre-test NGRT reading comprehension						

Results Table 16 - Exploratory analysis for Targeted Intervention – Model 18

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT Overall reading						
Group						
Pre-test NGRT reading accuracy						

Analyses of implementation factors

Results table 17 – Analysis of implementation factors for Whole Class Intervention – Model 19

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Pupil meta-cognition and comprehension behaviour						
Pupil Engagement						
Teacher Engagement						
Dosage						

Results table 18 – Analysis of implementation factors for Whole Class Intervention – Model 20

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	

Post-test NGRT reading comprehension						
Pupil Engagement						
Teacher Engagement						
Dosage						

Results table 19 – Analysis of implementation factors for Whole Class Intervention – Model 21

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT overall reading ability						
Pupil Engagement						
Teacher Engagement						
Dosage						

Results table 20 – Analysis of implementation factors for Targeted Intervention – Model 22

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Pupil meta-cognition and comprehension behaviour						
Pupil Engagement						
Teacher Engagement						
Dosage						

Results table 21 – Analysis of implementation factors for Targeted Intervention – Model 23

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT reading comprehension						
Pupil Engagement						
Teacher Engagement						
Dosage						

Results table 22 – Analysis of implementation factors for Targeted Intervention – Model 24

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	

Post-test NGRT overall reading ability						
Pupil Engagement						
Teacher Engagement						
Dosage						

Subgroup Analyses

Results Table 23 – Subgroup analysis for Whole Class Intervention – Model 25

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT overall reading ability						
Group						
Pre-test NGRT overall reading ability						
FSM						
SEN						

Results Table 24– Subgroup analysis for Whole Class Intervention – Model 26

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT reading comprehension						
Group						
Pre-test NGRT reading comprehension						
FSM						
SEN						

Results Table 23 – Subgroup analysis for Targeted Intervention – Model 27

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT overall reading ability						
Group						
Pre-test NGRT overall reading ability						
FSM						
SEN						

Results Table 24– Subgroup analysis for Targeted Intervention – Model 28

Model	Coefficient	S.E.	z	Sig.	95% Conf. Intervals	
Post-test NGRT reading comprehension						
Group						
Pre-test NGRT reading comprehension						
FSM						
SEN						

References

Armstrong, R.A., 2014. When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), pp.502-508.

Borman, G.D., Slavin, R.E., Cheung, A.C., Chamberlain, A.M., Madden, N.A. and Chambers, B., 2007. Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, *44*(3), pp.701-731.

Biggart, A., Kerr, K., O'Hare, L. and Connolly, P., 2013. A randomised control trial evaluation of a literacy after-school programme for struggling beginning readers. *International Journal of Educational Research*, *62*, pp.129-140.

Cox, D.R. and Reid, N., 2000. *The theory of the design of experiments*. CRC Press.

Granada Learning, 2013. https://www.gl-assessment.co.uk/sites/gl/files/images/New-Group-Reading-Test-NGRT_Data_Analysis.pdf

Torgerson, C.J. and Torgerson, D.J., 2007. The use of minimization to form comparison groups in educational research. *Educational Studies*, *33*(3), pp.333-337.

Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K. and Miller, D., 2011. Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement*, *22*(3), pp.265-289.

Wason, J. M., Stecher, L., & Mander, A. P. (2014). Correcting for multiple-testing in multi-arm trials: is it necessary and is it done?. Trials, 15(1), 364.

Appendix 1 – Stata Syntax for Primary Analysis

Universal Intervention

Model	*Dependent variable*	*Independent variable 1*	*Independent variable 2*	*Stata syntax*
1	Post-test – NGRT Overall reading score	Group – Control or Intervention	Pre-test – NGRT Overall reading score	xtmixed PostTestNGRTOverall Group PreTestNGRTOverall || School:
2	Post-test – NGRT Reading Comprehension	Group – Control or Intervention	Pre-test – NGRT Reading Comprehension	xtmixed PostTestNGRTComprehension Group PreTestNGRTComprehension || School:

Targeted Intervention

Model	*Dependent variable*	*Independent variable 1*	*Independent variable 2*	*Stata syntax*
3	Post-test – NGRT Overall reading score	Group – Control or Intervention	Pre-test – NGRT Overall reading score	xtmixed PostTestNGRTOverall Group PreTestNGRTOverall || School:
4	Post-test – NGRT Reading Comprehension	Group – Control or Intervention	Pre-test – NGRT Reading Comprehension	xtmixed PostTestNGRTComprehension Group PreTestNGRTComprehension || School: