Statistical Analysis Plan for Dialogic Teaching

Sheffield Hallam University

INTERVENTION	**Dialogic Teaching: Improving classroom talk**
DEVELOPER	CPRT/IEE
EVALUATOR	Sheffield Hallam University
TRIAL REGISTRATION NUMBER	ISRCTN14312500
TRIAL STATISTICIAN	Peter Thomas
TRIAL CHIEF INVESTIGATOR	Dr Tim Jay
SAP AUTHOR	Tim Jay, Peter Thomas, Sean Demack
SAP VERSION	2
SAP VERSION DATE	18th November 2016
EEF DATE OF APPROVAL	
DEVELOPER DATE OF APPROVAL	
SAP versions	1.0 - 18 November 2016 2.0 - current 9 December 2016 - edits following peer review

Introduction

The Dialogic Teaching intervention is designed to improve the quality of classroom talk as a means of increasing pupils' engagement, learning, and attainment, particularly those from disadvantaged backgrounds. The programme builds on the Dialogic Teaching approach previously developed by Professor Alexander, and successfully piloted in several local authorities. It emphasises dialogue through which pupils learn to reason, discuss, argue and explain, in order to develop their higher order thinking and articulacy.

The programme includes training for teachers, ongoing in-school monitoring and support, a pack containing study and reference materials, and a development and mentoring manual. Schools have been loaned all necessary equipment for video and audio recording, which are essential parts of the process, and cover has been paid for any time when participating teachers have been away from their classrooms. In the first year of the trial, training was provided to all teachers of pupils in Year 5 in intervention schools. At the end of the first year, teachers of Year 6 (i.e. the same cohort of pupils as received the intervention in the first year) in intervention schools will be offered the same training, reference materials, and equipment, although this will not be a compulsory part of the trial.

The main evaluation report will be published in autumn 2016, with an addendum including analysis of participating pupils' KS2 attainment to be completed in December 2017.

Study design

Research questions:

1. Does the Dialogic Teaching intervention improve pupil attainment across the curriculum after one year (science, maths, and literacy)?
2. Does the Dialogic Teaching intervention improve Key Stage 2 attainment in English and Mathematics after 2 years?

Schools in Leeds, Birmingham and Bradford were invited to participate in the trial. Recruitment was led by the Delivery Team (Robin Alexander from CPRT, supported by Mark Longmore from IEE).Eligible schools were those having at least two Year 5 classes, and a high proportion of FSM children (over 20%). The target number of schools was 80 (see sample size calculations below). 78 schools were recruited, with two pairs of these acting as federated schools; thus in this document we refer to 76 participating schools in total.

This trial employs a 3-level (pupils within classes within schools) Clustered RCT design. Randomisation is at school level, with half of the schools forming the intervention group and half of the schools forming a control group.

The primary outcome measures for the trial are measures of children's attainment in mathematics, English, and science. One-third of children (approximately 10 per class) were assessed in mathematics at the end of the intervention, one-third were assessed in English, and the remaining one-third were assessed in science[1].

Sample size calculations (Using 'Optimal Design'[2]):

- A power analysis for this 3-level CRT design, with 80 schools, 2 classes per school and 10 pupils per class has been carried out. We have assumed that the Intra-Cluster Correlation Coefficient (ICC) for both class and school levels is 0.1 and that Key Stage 1 attainment

[1] As the average class size is 30, each class was divided into 3 groups at random by the evaluation team. One group were assigned the maths test, one group the English test, and the third group took the science test.

[2] Raudenbush, S. W., et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]. Available from www.wtgrantfoundation.org.

scores (used as a baseline measure) account for 36% of the variance in the primary outcome measure (r=.6).

- The results of this analysis suggest an MDES (minimum detectable effect size) of approximately 0.25 with power of 0.8.
- Power calculations have also been carried out for this design for the subgroup of pupils eligible for FSM. Assuming 2 pupils eligible for FSM per class (~20%), and other assumptions as above, MDES is approximately 0.42.

Randomisation

Minimisation methods[3] were employed in order to achieve balance across intervention and control groups. The MinimPy[4] software package was used.

Minimisation was carried out at school level. Schools were allocated to the treatment or control group using the minimisation method, based on the percentage of pupils within each school eligible for free school meals (using the everFSM definition), the percentage of pupils using English as an additional language (EAL) and school-level KS2 total point score in 2013/14. For each variable entered into the minimisation procedure, a median split was used to designate schools as either 'high' or 'low' for that measure.

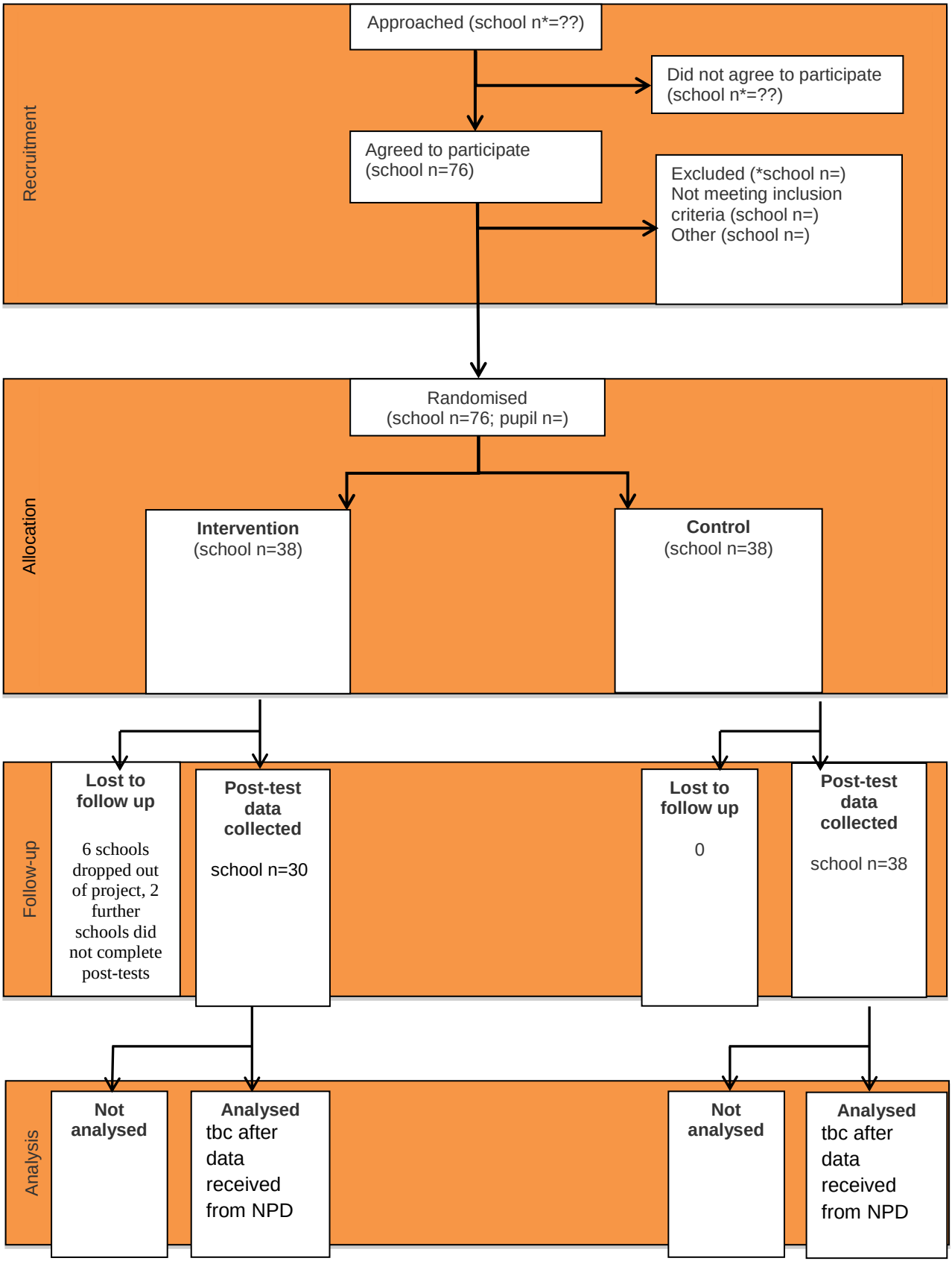
The minimisation produced a group of 38 intervention schools and 38 control schools.

Follow-up

[3] Pocock, S. J., & Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics, 103-115.

[4] Saghaei, M. and Saghaei, S. (2011) Implementation of an open-source customizable minimization program for allocation of patients to parallel groups in clinical trials. Journal of Biomedical Science and Engineering, 4, 734-739. doi: 10.4236/jbise.2011.411090. Available at: http://www.scirp.org/journal/PaperInformation.aspx?PaperID=8518

Participant flow diagram



* The delivery team led on recruitment and we have not been provided with this level of detail.

Outcome measures

Primary outcome

English attainment was measured using the Progress Test in English assessment (GL assessment). This test provides measures of reading accuracy, fluency and comprehension, standardised for UK populations. This test was administered as a pencil and paper test in groups. The main body of the test was administered, not including the sections on spelling, punctuation and grammar.

Mathematics attainment was measured using the Progress Test in Maths assessment (GL assessment). This test provides measures of fluency in facts and procedures, fluency in conceptual understanding, mathematical reasoning and problem solving, and has been standardised for UK populations. The main body of the test was administered, not including the section on mental arithmetic.

Science attainment was measured using the Progress Test in Science assessment (GL assessment). This test provides measures of knowledge, and ability to work scientifically, in physics, chemistry and biology, and has been standardised for UK populations. The whole test was administered.

For each measure, the raw, unstandardised, score will be used in analyses. This is a change from the protocol where we stated we would use age standardised scores, but this is not possible as we are using parts of some of the tests so the standardisation will not work and in addition we do not have dates of birth for the pupils.

Separate analyses will be carried out for each of the three primary outcome measures. For the intervention to be considered to have had an effect on attainment across the curriculum, there should be evidence of a positive effect on all three measures.

Secondary outcomes

Secondary outcomes will be Key Stage 2 point scores in English and Mathematics. These tests will be taken by pupils in May 2017, approximately two years after the start of the intervention. Data will be obtained from the National Pupil Database when they become available (unamended data are expected to be available in September 2017). However, it is understood that the effects of the intervention are likely to be mitigated, and possibly compromised, by (a) change of teacher (from a teacher trained in dialogic teaching to one who is not) from Y5 to Y6, (b) change in pedagogy as pupils approach the KS2 SATs (less extended dialogue, more IRE and text-based teaching).

Analysis

Primary intention-to-treat (ITT) analysis

The primary analysis of the impact of the intervention will be a multilevel linear regression model of each primary outcome measure. These models will have the primary outcome measure (mathematics, English, science attainment) as the dependent variable, with the following covariates included:

- School level (variables included in minimisation): Membership of the intervention group; percentage of pupils within each school eligible for free school meals (FSM); percentage of pupils using English as an additional language (EAL); school level average KS2 total point score 2013/14
- Pupil participant level: KS1 total point score (from NPD)
- Pupil participant level: FSM status

The modelling will be conducted in stages; first, a main effects stage followed by an interaction stage.

The main effects models will be used for the main results. These will be constructed in two steps:

1. Pupil participant variables: intervention / control dummy variable & KS1 attainment

2. As 1. but also including FSM status (pupil level) and the school level variables included in minimisation (%FSM; %EAL & School level KS2 attainment)

The second step represents the models that will be used for the main results for the three primary outcomes for this trial. The first step is included for sensitivity and so that the findings can be comparable across different trials.

The main effects model assumes that any impact of the intervention is consistent across different participant subgroups (for example FSM and non-FSM participants). To explore whether this is a reasonable assumption, interaction terms will be introduced one at a time.

The interaction terms to be included are:

- FSM*intervention (isolating FSM participants who received the intervention). These models explore whether the impact of the programme under evaluation depends upon FSM status. For example, the programme might have a greater impact for FSM participants (indicated by a positive coefficient on the interaction term) or it might have a greater impact amongst non-FSM participants (negative coefficient on the interaction term);
- KS1 point score*intervention (isolating the pre-test scores of intervention group participants). These models explore whether the impact of the programme under evaluation depends upon prior attainment (at KS1). For example, the programme might have a greater impact amongst higher attainers (indicated by a positive coefficient on the interaction term) or it might have a greater impact amongst lower attainers (negative coefficient on the interaction term).

If an interaction term is found to be statistically significant, a subgroup analysis will be used to explore this in more depth. i.e., if a FSM*intervention interaction term is found to be statistically significant, separate main effects models would be run for the FSM and non-FSM sub-samples to explore this; if a KS1*intervention term is significant, separate main effects models would be run on subsamples of pupils with differing levels of KS1 attainment (e.g. quartiles).

Analysis will be conducted using the 'mixed' command within Stata v13. An example of the STATA code is provided below

STATA v13 code for a 3-level multilevel linear regression model:

```
mixed [OutcomeVariable] InterventionDummy KS2(centred) FSM || School: || class:

estat icc
```

Imbalance at baseline

Baseline comparison of treatment and control groups: To evaluate whether random assignment resulted in equivalent groups at baseline, we will compare selected school-level baseline characteristics of the treatment and control group schools in the impact sample (see Table 4). Where appropriate we provide means and standard deviations of these baseline characteristics disaggregated by treatment and control group status.

Missing data

Missing data presents a problem for analysis, whether a pupil is missing a value for an outcome variable (post-test score) or for covariates (e.g. pre-test score). If outcome data is ‘missing at random’ given a set of covariates then the analysis has reduced power to detect an effect; if data is ‘missing not at random’ (for example, differential dropout in the intervention and control groups for unobserved reasons) then omitting these pupils (as in the primary ‘completers’ analysis) could bias the results. Imputing missing data could improve the robustness of the analysis and examine how sensitive the results are to alternative assumptions.

Eight schools from the intervention group did not return primary outcome measures data. This could potentially bias the results of the primary analysis. Comparisons between the thirty intervention

schools that returned data, and the 38 control schools, will be carried out using t-tests for school-level variables including the proportion of pupils eligible for FSM, and 2013/14 KS2 attainment; this will provide a check on balance at post-test.

If the amount of missing data at pupil-level within schools exceeds 5% of the randomised sample then we will consider using multilevel multiple imputation methods as a sensitivity analysis to check whether the results from the 'completers' analysis are sensitive to missing data. It may be that the results of the multiple imputation do not differ appreciatively from the 'completers' analysis. If this is the case and we are reasonably confident that covariates explain any 'missingness' then this will complete the primary analysis.

Secondary outcome analyses

An addendum to the main report will be completed in December 2017, including analysis of KS2 attainment.

At the end of the second year, we will again employ the 3-level clustered design, with KS2 point scores for English reading and Mathematics reasoning as outcome variables, KS1 scores as covariate, and experimental group and eligibility for free school meals as predictors. Key Stage 2 attainment tests will be taken by pupils in May 2017, approximately two years after the start of the intervention. Data will be obtained from the National Pupil Database when they become available (un-amended data are expected to be available in September 2017).

For the secondary outcomes, data on all 38 intervention and 38 control schools will be obtained from the NPD.

On-treatment analysis

All schools that completed and returned primary outcome measures participated fully in the intervention. Therefore analysis of primary outcome measures as described above will in itself constitute an on-treatment analysis (the schools that completed the testing are those schools that fully participated in the intervention).

For the secondary outcome analysis, a main analysis will be conducted that includes all pupils in all schools at the time of the project, and an on-treatment analysis will be conducted that includes in the intervention group only those 30 schools that participated fully in the intervention for the two terms of the project.

Subgroup analyses

As noted above, subgroup analyses of pupils eligible for FSM will be carried out for both primary and secondary analyses. Tests for interaction will be conducted to assess whether or not there are differential effects for pupils eligible for FSM relative to other pupils.

Subgroup analyses of pupils with low prior attainment will be carried out for both primary and secondary analyses. Tests for interaction will be conducted to assess whether or not there are differential outcomes for children with low prior attainment relative to other pupils.

If an interaction term is found to be statistically significant, a subgroup analysis will be used to explore this in more depth. For example, if a FSM*intervention interaction term is found to be statistically significant, separate models would be run for the FSM and non-FSM sub-samples to explore this. Although not explicitly specified in the protocol, analysis would be conducted on the non-FSM sub-sample as well as the FSM sub-sample, as previous research has found the non-FSM sub-sample to be statistically significant (e.g. Maxwell et al 2014).

Effect size calculation

In multilevel models or mixed effect models we assume that variations in a post-test outcome are due to different sources, which must be fully accounted for in a statistical model. For cluster randomised

trials, the total variability can be decomposed into random variation between pupils (σerror2) and heterogeneity between classes and schools (σS2). Effect sizes for cluster randomised trial using total variance can be calculated as:

$$ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\sigma_S^2 + \sigma_{error}^2}}$$

Where (ȲT−ȲC) adjusted denotes ANCOVA mean difference between intervention groups adjusting for pre-test score and the variables used within the minimisation..

Tables

Key conclusions
1. Headline for schools 1
2. Headline for schools 2
3. Important factors for implementation 1
4. Important factors for implementation 2
5. Possible further research question

Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	EEF cost rating
Treatment 1 vs. control (GL Progress Test English)				
Treatment 2 vs. control (GL Progress Test Mathematics)				
Treatment 3 vs. control (GL Progress Test Science)				
Treatment 1 FSM vs. control				
Treatment 2 FSM vs. control				
Treatment 3 FSM vs. control				
Treatment 1 non-FSM vs. control				
Treatment 2 non-FSM vs. control				

Treatment 3 non-FSM vs. control				

Table 2: Timeline

Date	Activity

Table 3: Minimum detectable effect size at different stages

Stage	N [schools] (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	80 (40; 40)	0.6	0.10	minimisation	80%	0.05	0.25
Randomisation	76 (38;38)	0.6	0.10	minimisation on FSM, EAL and KS2	80%	0.05	0.26
Analysis (i.e. available pre- and post-test)							

Example Table 4: Baseline comparison

Variable	Intervention group		Control group	
School-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for FSM				
English as additional language (EAL)				
School type				
Ofsted rating	…	…	…	…
Pupils with SEN				
School-level (continuous)	**n (missing)**	**[Mean or median]**	**n (missing)**	**[Mean or median]**
Number of Y5 pupils				
	…	…	…	…
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for FSM				
English as additional language (EAL)				
Gender	…	…	…	…
Pupil-level (continuous)	**n (missing)**	**[Mean or median]**	**n (missing)**	**[Mean or median]**
Pre-test score				

Example Table 5: Primary 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
Post-test (GL Progress Test English)							
Post-test (GL Progress Test Mathematics)							
Post-test (GL Progress Test Science)	…	…	…	…	…	…	…

Effect sizes for cluster randomised trial using total variance are calculated as noted above. Confidence Interval calculations are as noted above.

Conclusion

Key conclusions
1.
2.
3.
4.
5.

References

What Works Clearinghouse (2008). *Procedures and standards handbook* (Version 2.0). *Retrieved June*, *6*, 2011

Hedges, L. V. (2007). 'Effect sizes in cluster-randomized designs.' *Journal of Educational and Behavioral Statistics*, *32*(4), 341-370.

Hutchison, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. Slough, UK: NFER.

Maxwell, B., Connolly, P., Demack, S., O'Hare, L., Stevens, A., & Clague, L. (2014). *TextNow Transition Programme*.

Torgerson, D. J., & Torgerson, C. J. (2008). *Designing randomised trials in health, education and the social sciences: an introduction*. Palgrave Macmillan.