

Statistical Analysis Plan for multi-arm research dissemination trial: Passive Trial.

NFER



INTERVENTION	Multi-arm research dissemination trial: Passive Trial.
DEVELOPER	Four organisations: The Centre for Evaluation and Monitoring (CEM) at Durham University; Train Visual with Campaign for Learning; The Institute for Effective Education (IEE) at the University of York; and NatCEN (working on behalf of ResearchEd).
EVALUATOR	National Foundation for Educational Research
TRIAL REGISTRATION NUMBER	ISRCTN84508118
TRIAL STATISTICIAN	Adam Rabiasz
TRIAL CHIEF INVESTIGATOR	Ben Styles
SAP AUTHOR	Adam Rabiasz and Ben Styles
SAP VERSION	2
SAP VERSION DATE	27 th January 2017
EEF DATE OF APPROVAL	
DEVELOPER DATE OF APPROVAL	

Table of contents

Table of contents.....	1
Introduction.....	3
Study design	3
Protocol changes	4
Randomisation	4
Calculation of sample size	4
Outcome measures.....	4
Primary outcome.....	4
Secondary outcome	4
Analysis	4
Primary intention-to-treat (ITT) analysis	5
Imbalance at baseline	5
Missing data.....	5
On-treatment analysis.....	6
Secondary outcome analysis.....	6
RESTRICTED (PUBLIC when signed off)	

Subgroup analyses	6
Effect size calculation	6
Report tables	6
References	6

Introduction

The Education Endowment Foundation (EEF) has commissioned a number of providers to use a range of different methods of communicating research findings and evidence to teachers and schools. EEF wishes to determine the impact that these different approaches have on pupils' literacy.

The primary research question is:

- what are the effects of different ways of communicating research evidence and findings to teachers and schools, and different ways of engaging them with research, on pupil attainment?

NFER and the providers have attempted to answer this question by running, in parallel, two multi-armed RCTs; one which will assess the impact of passive means of communication and a second which will assess the impact of active and passive means of communication (see the Active Trial SAP). A difference between the passive and active trials is that schools will not be aware they are part of the passive trial so there was no recruitment phase for this trial. In the passive trial, schools were randomised to one of five groups: the intervention groups where providers will use 'passive' means of communicating with schools or to a 'business as usual' control group.

To determine the impact that different approaches to communicating research evidence has on attainment, NFER will measure the development of pupil literacy using Key Stage 1 attainment as a baseline and Key Stage 2 attainment as the outcome measure.

Study design

The sampling frame of the passive trial was defined to include all state primary schools in England that have a Year 6, that have not been recruited to the active trial and which are not already in receipt of research communications from the providers. From this population a sample of 12,500 schools was drawn and split into five equal-sized groups; one for the passive arm of each provider and one control group of the trial. After randomisation the providers were given their list of schools and proceeded to implement their passive communication methods. Note that schools which were included in the passive trial were not informed of their participation in the trial.

Description of the communication methods used in the passive trial that each provider tested are outlined below:

1. **The Institute for Effective Education at York University** looked at the impact of:

IEE's Better magazine, Best Evidence in Brief publication and Evidence4Impact website. This will entail: Printed and electronic materials that explain research findings and identify effective interventions. The IEE already produces research summaries as part of its remit to improve the use of evidence in the profession, including magazines that bring together evidence on particular topics (for example, assessment, or struggling readers) and fortnightly emails summarising new research findings. It launched a website shortly before the trial protocol was written, Evidence 4 Impact, providing a searchable database of evidence-based education programmes.

2. **Teaching How2s**, in partnership with Campaign for Learning (an existing EEF grantee), compared:

Invitation to use Teaching How2s. Inviting schools to subscribe to a school log-in to the Teaching How2s website. This provides a large number of visual guides to evidence-based teaching techniques. These are carefully designed, step-by-step presentations that walk teachers through activities that they can use in their classroom. The idea is that visual guides can accurately communicate evidence-based teaching strategies to teachers and help implement them in practice.

3. **Centre for Evaluation and Monitoring**, part of Durham University:

CEM booklet: Sending out a handy, up-to-date, easy-to-read booklet on research-based strategies for teaching literacy in Key Stage 2, 'tips for teachers' cards and regular posters to foster engagement

4. **ResearchEd and NatCen** worked together to deliver:

Invitation: Teachers were invited to attend a conference on current research in literacy. The speaker list was largely comprised of academics, researchers and associated school representatives.

Protocol changes

p.5 – Section 3.3 – outcome measures to use amended NPD data available in December 2016 and December 2017 (rather than unamended data in September 2016 and September 2017)

Randomisation

Simple randomisation was used to randomly allocate 12,500 schools to 1 of 4 passive arms or a control group (all groups had 2,500 schools). Once randomisation was completed providers received their list of schools and implemented their passive communications strategies with them.

Calculation of sample size

Power calculations performed in the protocol indicated that with four passive arms and a control group, each containing 2,500 schools, the estimated MDES would be 0.024. Sample size calculations were based on Key Stage 2 outcomes with Key Stage 1 as a baseline using the following assumptions: average of 34 pupils per cohort per school; ICC=0.15 (reduced from 0.2 through the use of Key Stage 1 as a covariate); correlation between Key Stage 1 and Key Stage 2=0.7; power=80% and significance=5%.

Outcome measures

Primary outcome

The primary outcome is 2016 Key Stage 2 attainment in English. Our preference is to use raw test scores as opposed to scaled scores since some information is lost in the scaling process¹. However, it appears from the October 2016 version of NPD tables that raw scores will not be available. The primary outcome will therefore be the sum of READSCORE (scaled score in reading) and GPSSCORE (scaled score in Grammar, Punctuation and Spelling). Writing is not included as this was teacher assessed and is therefore vulnerable to bias.

Secondary outcome

Key Stage 2 results from 2017.

Analysis

The analysis will follow EEF guidelines².

¹ <https://www.gov.uk/guidance/scaled-scores-at-key-stage-2>

² https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analyses_for_EEF_evaluations.pdf

Primary intention-to-treat (ITT) analysis

The primary outcome analysis will be 'intention to treat'. A multilevel model with two levels (school and pupil) will be used for the analysis to account for the cluster randomisation. The first step will be to determine if different methods of communicating research has had any significant effect on Key Stage 2 attainment. This will be ascertained by fitting two models to the data: one which has Key Stage 2 attainment as the dependent variable and Key Stage 1 attainment, as measured by the sum of the NPD variables KS2_KS1READPS (Key Stage 1 reading point score) and KS2_KS1WRITPS (Key Stage 1 writing point score); and a second model identical to the first but including 4 dummy variables that indicate which group an individual is in (the default being control). For each model, we will obtain the value of the likelihood, L . The likelihood ratio test statistic is computed as:

$$-2\log L_1 - (-2\log L_2)$$

which under the null hypothesis follows a chi-squared distribution on 4 degrees of freedom since 4 parameters were added to create the second model. The null hypothesis will be rejected if $p < 0.05$ for this chi-squared test. This will provide a global test for the impact of the passive communication strategies used within this trial.

Analysis will continue in terms of presenting effect sizes and confidence intervals for each arm versus control, regardless of the outcome of the global test. No *post hoc* tests will be made between the trial's intervention arms (which was a possibility in the original protocol) as the emphasis of the study is to determine the effectiveness of each individual approach to passive communication.

The issue of multiple comparisons (type I error in particular) must be considered when calculating confidence intervals. Methods used to control for the family-wise error rate can be classified into one of two types; step-down tests and single-step tests. Step-down tests are more powerful than single-step tests but they are unable to produce confidence intervals, therefore since a confidence interval is required from this experiment attention shall focus on single-step tests.

Well known single-step tests such as the Bonferroni correction or Scheffé's method are known to be conservative. Tukey's range test (also known as the Tukey-Kramer method) is considered optimal in a broad range of circumstances. However, like a number of other approaches used to account for multiple comparisons it corrects for the family-wise error rate based on the assumption that all pairwise comparisons of the group means are being made. This multi-arm trial follows a many-to-one comparison scheme (i.e. each group mean is compared to the control group mean and only the control group mean) and therefore those methods that assume all pairwise comparisons in their calculations are too conservative.

Taking into account these considerations the two-sided form of Dunnett's Test (1955) will be used to control the family-wise error rate. This test is specifically designed to handle the scenario of multiple comparisons being made to a common control group, it is an exact test (i.e. its family-wise error rate is exactly equal to alpha) and can be used in balanced and unbalanced designs.

All multilevel analyses will be carried using the R package nlme. Dunnett's Test will be used to control the family-wise error rate using the R package multcomp

Imbalance at baseline

As the primary outcome is available from administrative data and there was no opportunity for schools to withdraw, it is anticipated that the level of missing will not exceed 5% at either the school or pupil level so no analysis of imbalance is planned.

Missing data

As this analysis uses administrative data, it is anticipated that the number of pupils missing will be very small and so these cases can be excluded from the analysis without risk of bias. It is anticipated that the level of missing will not exceed 5% at either the school or pupil level so no missing data analysis is planned. If it does exceed 5%, reasons for missingness will be explored. They are highly

likely to be administrative e.g. problems matching our list of schools to NPD due to the Unique Reference Number changes brought about by academisation. For data to be missing for a biased reason, a passive communication strategy would have to have interfered with a school's ability to administer Key Stage 2 tests. This seems entirely unlikely.

On-treatment analysis

n/a

Secondary outcome analysis

Follow-up outcomes using the Key Stage 2 English attainment data for the 2017 cohort will be analysed. The same analysis as that outlined in the primary intention-to-treat analysis for the 2016 cohort will be used.

Subgroup analyses

Sub-group analysis on the primary outcome will be carried out on the following group only as per the protocol and the most recent EEF analysis guidelines: whether a pupil has ever received free school meals (everFSM). This will be done using a model identical to the primary outcome model but including everFSM, and everFSM*intervention as covariates too. Analysis shall proceed as per the original primary outcome modelling i.e. the first model shall be identical to the primary outcome model but with everFSM as a covariate. The second model shall contain a further 8 covariates: 4 dummies for randomised group and 4 interaction terms.

A second model will also be run including only everFSM children, as per EEF analysis guidelines. Analysis shall proceed as per the original primary outcome model.

Effect size calculation

All effect sizes will be calculated using total variance from a multilevel model, without covariates, as the denominator i.e. equivalent to Hedges' g. Confidence intervals for each effect size will be derived from Dunnett's Test (see above) that takes into account multiple comparisons. Dunnett's Test will apply to the model coefficients themselves. These will be converted to effect size confidence intervals using the same formula as the effect size itself.

Report tables

The EEF trial report template³ contains several tables whose structure is pre-specified.

References

Dunnett C. W. (1955.) 'A multiple comparison procedure for comparing several treatments with a control' Journal of the American Statistical Association, 50:1096–1121.

³ <https://educationendowmentfoundation.org.uk/evaluation/resources-centre/writing-a-research-report/>