

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

NFER



INTERVENTION	Multi-arm research dissemination trial: Active 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 AUTHORS	Adam Rabiasz, Ben Styles and Pippa Lord
SAP VERSION	3
SAP VERSION DATE	3 rd February 2017
EEF DATE OF APPROVAL	
DEVELOPER DATE OF APPROVAL	

Table of contents

Table of contents.....	1
Introduction.....	3
Study design	3
Protocol changes	5
Randomisation	5
Calculation of sample size	5
Outcome measures.....	5
Primary outcome.....	5
Secondary outcomes	5
Analysis.....	6
Primary intention-to-treat (ITT) analysis	6
Imbalance at baseline	7
Missing data.....	7
CACE analysis	8
Secondary outcome analyses - attainment.....	8
RESTRICTED (PUBLIC when signed off)	

Secondary outcome analyses - research use	8
Path analysis.....	9
Subgroup analyses	9
Effect size calculation	9
Report tables	9
References	9
Appendix A: Literacy Octopus Engagement Categories.....	11

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 active means of communication, and a second which will assess the impact of passive means of communication. This SAP outlines the approach being adopted for the active trial, for further details on the passive trial see the relevant SAP.

The active trial schools were randomly allocated to one of ten groups. These ten groups comprised an active and passive arm for each of three of the providers, two active and one passive arms for the remaining provider, and a 'business as usual' control group. The active trial will also investigate the following secondary research questions:

- What is the effectiveness of different communication approaches in terms of schools' dispositions towards research?
- What is the degree to which research information is informing teaching and learning?

To determine the impact that the different approaches to communicating research evidence and findings 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.

To answer the secondary research questions, teacher responses to a standard baseline and outcomes survey that has been validated to measure 'Research Use' (Poet *et al.* 2015, unpublished) will be used to construct measures. The link between these measures and attainment will be analysed in a path analysis.

The aim of the analysis is to measure the differential progress in literacy development between pupils in each arm of the trial in order to measure the effect different approaches to disseminating research evidence and findings have on literacy attainment over three Year 6 cohorts (those of 2014/15, 2015/16 and 2016/17 respectively).

Study design

The main sample consisted of English primary schools with a Year 6 excluding special schools, independent schools and local authorities that were classed as too far from York and London for travel within one day (which would be required for one of the arms of the intervention). In addition, schools were excluded that were already taking part in work with the providers IEE and NatCen or research engagement projects with the funder EEF, the IoE or the National College's Evidence Based Teaching projects, teaching schools alliances.

An initial sample of 7,884 schools followed by a top-up sample of 2,922 schools were approached. Each arm of the trial required 60 recruited schools and the control group required 240 schools – a planned total of 780 schools. Once this target had been met, and indeed exceeded (823 schools were recruited), schools were randomly assigned to one of the nine intervention arms of the trial or to the control group (which was expanded in size to 283). After randomisation the providers were given a list of their schools and each proceeded to implement their communication methods (outlined below).

Descriptions of the active and passive communication methods that each provider tested are outlined below:

1. **The Institute for Effective Education at York University** looked at the impact of:
 - a. *Passive arm:* 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.
 - b. *Active arm:* An evidence fair that gives schools an opportunity to discover more about interventions that can help primary school pupils with their literacy. Senior leaders from schools will attend the events, and meet with proponents of evidence-based programmes and hear from other schools that have used them. There will also be pre and post conference support. Schools will also receive the materials from the passive arm.
2. **Teaching How2s**, in partnership with Campaign for Learning (an existing EEF grantee), compared:
 - a. *Passive arm:* Giving schools log-ins 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.
 - b. *Active arm:* Schools received access to the website, but also received additional support on how to use the How2s as part of their school improvement and teacher training plans. This included advice on how to use its communication and management tools to allow a school to track the use of the different guides, and gave teachers an opportunity to learn from others' use of the guides. Participating schools received an initial face to face induction, and ongoing emails suggesting how to effectively use the resources.
3. **Centre for Evaluation and Monitoring**, part of Durham University, wanted to compare:
 - a. *Passive arm:* 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
 - b. *Active arms:* Following up the booklet with various levels of training. One group will receive a light touch 'twilight' CPD session that reviews the material included in the leaflet. Another will receive further CPD session, diagnostic tests to identify individual pupils' literacy issues and support to learn from their fellow KS2 teachers through structured peer observations. These two groups will be treated as separate arms within the trial.
4. **ResearchEd and NatCen** worked together to deliver:
 - a. *Passive arm:* 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.

- b. *Active arm:* Alongside the invitation to the conference, attendees were invited to become part of an online community offering support and activities before and after the conference in order to help them to use research findings in their own schools.

Protocol changes

p.10 – Section 3.4 – use of amended Key Stage 2 data available in December 2016 (rather than unamended Key Stage 2 data available in September 2016). This change was made after advice from DfE concerning the data obtained from the new Key Stage 2 tests.

p.10 – Section 3.4 – incorporation of 2017 attainment data into a repeated-measures multi-level model. As this model can only be regarded as repeated measures at the school level it is debatable how useful it would be, so it is omitted from the SAP.

Randomisation

In the active trial 823 schools were randomly allocated to either one of the trial's nine arms (60 schools per arm) or the control group (283 schools) via school-level stratified randomisation. The stratum used was geographical area: a binary variable that indicated whether a school was situated in the north or south of England. School-level randomisation was used due to many of the communication strategies being evaluated involving whole-school 'mobilisation'.

Once randomisation was complete each provider received their list of schools and implemented their active communications methods with them.

Calculation of sample size

The power calculations performed in the protocol indicated that with nine arms, each having 60 schools and a control group of 240 schools the estimated minimum detectable effect size (MDES) of the active trial would be 0.121. Assumptions underpinning this calculation were: an average cohort size of 36 pupils per school, ICC=0.15 (reduced from 0.2 through the use of Key Stage 1 attainment data as a covariate), correlation of 0.7 between Key Stage 1 performance and Key Stage 2 performance, 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 outcomes

In addition to the Key Stage 2 results of the 2016 cohort, the Key Stage 2 results from 2015 and 2017 will also be analysed. 2015 was the last year of the old Key Stage 2 tests and this outcome will therefore be KS2_ENGTOTMRK; total marks achieved in English test (sum of reading and writing tests). The 2015 data will be discussed in the main report in 2017. For 2017 data the outcome will be the same as for 2016 and will be reported in an addendum report in 2018.

Other secondary outcomes that the trial will study are degree of research use: awareness, understanding and action. These will be measured by calculating the factor scores based on the

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

teacher responses to the standard Research Use outcomes survey (Poet *et al.* 2015, unpublished) administered in January 2016. Each teacher's score for a given factor is formed by summing together the responses of the survey items that have loaded onto that factor. The following six factors are measured by the research use outcomes survey:

- Positive disposition to academic research in informing teaching practice (Measure 1)
- Uses academic research to inform selection of teaching approaches (Measure 2)
- Perception that academic research is not useful in learning (Measure 3)
- Perception that own school does not encourage use of academic research (Measure 4)
- Active engagement with online evidence platforms (Measure 5)
- Your knowledge about research (Measure 6)

Note that the research use outcomes survey was in development whilst the trial protocol was being written (i.e. the six measures were not yet fully specified). Hence, at the time of writing the protocol, we anticipated that the secondary outcome questions could cover: the effectiveness of the different communication approaches in terms of schools' dispositions towards research, and the degree to which research information is informing teaching and learning. Measures 1 and 2 most clearly meet our suggested outcome questions but Measures 3 – 6 will also be useful to explore as secondary outcomes. See 'Secondary outcomes analysis – research use' section below for more detail.

Analysis

The analysis will follow EEF guidelines².

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), in addition to a dummy variable identifying if the school is in the north to account for the stratified randomisation; and a second model identical to the first but including 9 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 9 degrees of freedom since 9 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 as the emphasis of the study is to determine the effectiveness of each individual approach to passive communication.

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

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, it is anticipated that the level of missing will not exceed 5% at either the school or pupil level. As the NPD data used in this trial is de-identified, it is obtained irrespective of school withdrawal. No analysis of imbalance is therefore 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 communication strategy would have to have interfered with a school's ability to administer Key Stage 2 tests. This seems entirely unlikely.

In the event that more than 5% of cases are found to be missing for a possibly biased reason, some further analysis will be carried out. In particular, a logistic multilevel model of whether or not an individual is missing, regressed on baseline test score, stratification and randomised group dummies, will help determine the extent of bias.

Missing data generally 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 with 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. It can also signal missing not at random if the imputed result is much different from completers analysis. Likelihood based methods (e.g. nlme function in R) are usually consistent with the results from multiple imputation if the missingness mechanism is missing at random. If it is not, some sensitivity analysis, for example using extreme values, may be necessary.

CACE analysis

Fidelity analysis will be conducted on individual models containing data from each arm, in turn, with control schools i.e. nine models. A further variable that measures each school's level of engagement with trial activity will be matched with the attainment data. This additional variable will measure four levels of engagement: none, low, mid and high. Whilst the quantitative and qualitative characteristics of each level of engagement are specific to the trial arm, we adopted an overall consistent approach to categorisation, namely:

Level of engagement	Description of engagement according to Monitoring Information (MI) data
None	No engagement
Low	Engaged less than expected for that arm
Mid	Engaged as expected for that arm
High	Engaged more than expected for that arm

Each provider adopted a pre-agree MI data tool, with pre-specified fields in order to capture the qualitative and quantitative data required. Providers submitted this data at three time-points (April 2015, July 2015 and April 2016). NFER researchers carried out the categorisation. Appendix A provides the detailed engagement schemes applied to each arm.

In order to obtain a more accurate measure of the 'pure' dosage effect of each communication method on attainment the CACE impact estimate will be calculated. Because schools may potentially have unobserved characteristics that have an influence on both compliance with the trial and academic attainment a two stage least squares model will be used to calculate the CACE estimate (Angrist and Imbens, 1995).

The first stage of the model will be engagement level regressed on all covariates that are used in the main primary outcome model and in addition will include, as an instrumental variable, a binary variable that indicates a pupil's pre-intervention treatment allocation. The second stage of the model will regress the primary outcome on the covariates used in the main model and will also include a covariate representing the pupil's estimated dosage level from the first stage of the model and an interaction term between the estimated dosage and the pupil's pre-intervention treatment allocation. The coefficient of the interaction term is the CACE estimate of the dosage effect. In the event that there are no confounding factors affecting compliance and attainment the CACE estimate will be equal to the intention-to-treat estimate.

A further factor that must be taken into account is the hierarchical nature of the data. To ensure that this factor of the data is accounted for correctly the R package ivpack, which has the functionality to correctly handle hierarchical data when using instrumental variables, will be used to perform the CACE analysis.

Secondary outcome analyses - attainment

Interim and follow-up outcomes using the Key Stage 2 English attainment data for the 2015 and 2017 cohorts will be analysed. The same analysis as that outlined in the primary intention-to-treat analysis for the 2016 cohort will be applied to both cohorts.

Secondary outcome analyses - research use

All models using research use data will be affected by the substantial levels of attrition experienced at follow-up survey administration. Each model will therefore be proceeded by a multilevel logistic model consisting of two levels (school and teacher) of whether or not an individual is missing at follow-up, regressed on the covariates of the main model. This will help to determine the extent of bias. As some teachers at follow-up completed the survey who did not complete one at baseline, multilevel multiple

imputation may be useful. The completers model described in the next paragraph will be compared with one following imputation of the baseline measure.

Each of measures 1 and 2 from the research use survey will be the dependent variable in a separate multilevel model containing two levels (teacher and school). Note that as we shall use this model only to estimate coefficients and their standard error, the variable cluster size apparent during the teacher survey should not be a problem assuming missing at random (Snijders, 2005). The covariates that will be entered into each model will be the equivalent baseline research use measure and a dummy variable to identify schools in the north, as per the stratified randomisation. Analysis will proceed as for the primary outcome.

The remaining measures 3 – 6 will be analysed at the school-level using ANOVA on follow-up data only.

Path analysis

Path analysis will be performed in an attempt to uncover the link between the extent of research use and attainment via a multilevel model containing two levels (school and teacher). Each of the six measures of research use at follow-up will be entered into the model as teacher-level covariates along with the school mean for Key Stage 1 baseline. The dependent variable will be the school mean of the primary outcome variable. The model will be checked for tolerance as the research-use measures are likely to be correlated.

Subgroup analyses

Sub-group analysis on the primary outcome will be carried out on the following group only as per the protocol for active trial 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 18 covariates: 9 dummies for randomised group and 9 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

Angrist, J.D., & Imbens, G.W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430), 431-442.

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/>

Poet, H., Mehta, P. and Nelson, J. (2015). 'Research Use in Schools - Survey, Analysis and Guidance for Evaluators.' Slough: NFER.

Snijders, Tom A.B. (2005) 'Power and Sample Size in Multilevel Linear Models.' In: B.S. Everitt and D.C. Howell (eds.), *Encyclopedia of Statistics in Behavioral Science*. Volume 3, 1570–1573. Chichester (etc.): Wiley, 2005.

Appendix A: Literacy Octopus Engagement Categories

NFER researchers carried out the engagement categorisation at school level. Some qualitative judgements were made by NFER coders where quantitative data was borderline in the categories below. Some double-coding was undertaken on the initial datasets in April 2015, to quality assure the consistency of applying the coding schemes. The detailed engagement schemes were as follows:

Arm 1: IEE Passive	
Level of engagement	Description
None	0% materials opened (incl. bounced back after several attempts / unable to send)
Low	At least one teacher opened a small amount of materials (e.g. 1 – 20% of materials)
Mid	At least one teacher opened a some materials (e.g. 21-50% materials)
High	At least one teacher opened most materials (e.g. more than 50% of materials)

NB – MI data for this arm was collated at an individual teacher level. NFER applied the above aggregated metric at school level (i.e. one engagement code per school).

Arm 2: IEE Active	
Level of engagement	Description
None	Did not attend Evidence Fair; 0% materials opened (incl. bounced back after several attempts / unable to send)
Low	Did not attend Evidence Fair; at least one teacher opened a small amount of material (e.g. 1 – 20% of materials)
Mid	At least one teacher attended Evidence Fair; and at least one teacher opened some materials (e.g. 21-50% of materials)
High	More than one teacher attended Evidence Fair (NB – where only one teacher was listed, one teacher attended Evidence Fair); and at least one teacher opened most materials (e.g. more than 50% of materials)

NB – MI data for this arm was collated at an individual teacher level. NFER applied the above aggregated metric at school level (i.e. one engagement code per school).

Arm 3: Teaching How2s Passive	
Level of engagement	Description
None	No activity on How2s site (i.e. zero techniques and/or nudges recorded)
Low	Small number of active users (e.g. 1 – 3), and limited activity on How2s site (e.g. 1 – 15 techniques recorded)
Mid	Small number of active users (e.g. 2 – 5), and moderate activity on How2s site (e.g. 15 – 40 techniques recorded); or larger number of active users (i.e. 6 – 25 users) and limited/moderate activity on How2s site (e.g. 10 – 20 techniques recorded)
High	Small number of active users (e.g. 2 – 5), and high activity on How2s site (e.g. over 40 techniques recorded, and in some cases nudges recorded also); or large number of active users (e.g. 10+) and high activity on How2s site (e.g. over 40 techniques recorded, and in some cases nudges recorded also)

Arm 4: Teaching How2s Active	
Level of engagement	Description
None	Did not attend induction; no activity on How2s site (i.e. zero techniques and/or nudges recorded)
Low	Did not attend induction; small number of active users (e.g. 1 – 3), and limited activity on How2s site (e.g. 1 – 15 techniques recorded)
Mid	Attended induction; small number of active users (e.g. 2 – 5), and moderate activity on How2s site (e.g. 15 – 40 techniques recorded); or larger number of active users (i.e. 6 – 25 users) and limited/moderate activity on How2s site (e.g. 10 – 20 techniques recorded)
High	Attended induction; small number of active users (e.g. 2 – 5), and high activity on How2s site (e.g. over 40 techniques recorded, and in some cases nudges recorded also); or large number of active users (e.g. 10+) and high activity on How2s site (e.g. over 40 techniques recorded, and in some cases nudges recorded also)

Arm 5: CEM Passive	
Level of engagement	Description
None	Did not respond to initial contact, did not engage with materials (none downloaded)
Low	Responded to initial contact, did not engage with materials and did not ask for further materials
Mid	Responded to initial contact, engaged with materials
High	Responded to initial contact, engaged with materials, and engaged further proactively with provider

Arm 6: CEM Active Light	
Level of engagement	Description
None	Did not respond to initial contact, and no engagement with materials; or engaged with initial contact, but no engagement with materials
Low	Responded to initial contact, and engaged with materials, did not attend twilight CPD
Mid	Responded to initial contact, engaged with materials, and attended twilight CPD
High	Responded to initial contact, engaged with materials, attended twilight CPD, and engaged further proactively with provider

Arm 7: CEM Active	
Level of engagement	Description
None	Did not respond to initial contact, and no engagement with materials; or engaged with initial contact, but no engagement with materials
Low	Responded to initial contact, and engaged with materials, and attended the first round of twilight CPD only (i.e. CPD1)
Mid	Responded to initial contact, engaged with materials, and attended two of the support activities (i.e. from CPD1, CPD2, INCAS support)
High	Responded to initial contact, engaged with materials, and engaged in all three support activities (i.e. CPD1, CPD2, and INCAS). A special note was applied if the school also hosted a CPD session to indicate high levels of engagement.

Arm 8: ResearchED Passive	
Level of engagement	Description
None	School did not attend conference
Low	School did not attend conference
Mid	One teacher attended conference
High	More than one teacher attended conference

Arm 9: ResearchED Active	
Level of engagement	Description
None	School did not attend conference
Low	One teacher attended conference
Mid	More than one teacher attended conference; or one teacher attended conference, and school attended webinar conference
High	More than one teacher attended conference, and school attended webinar conference

The NFER team used the MI engagement data to help identify case studies. In addition, we assessed any change in engagement over time in Literacy Octopus activity, and applied a qualitative schema. We used this information to further help identify case studies.

Change in/continued engagement (2016 data compared qualitatively with 2015 data)
School did not engage in 2015, and did not engage in 2016
School engaged less than expected in 2015, and in 2016 had no further or limited engagement
School engaged as expected for their arm in 2015, and in 2016 this continued; OR, school had limited engagement in 2015, but in 2016 increased their engagement moderately
School engaged more than expected for their arm in 2015, and in 2016 continued to engaged highly; OR, school engaged moderately in 2015 and in 2016 increased their engagement