

Statistical Analysis Plan for Catch Up[®] Literacy: A follow up evaluation

Evaluating Institution National Foundation for Educational Research (NFER)



INTERVENTION	Catch Up Literacy – a follow up evaluation
DEVELOPER	Catch Up [®] 1
EVALUATOR	National Foundation for Educational Research (NFER)
TRIAL REGISTRATION NUMBER	ISRCTN11318637
TRIAL STATISTICIAN	Palak Roy
TRIAL CHIEF INVESTIGATOR	Simon Rutt
SAP AUTHOR	Palak Roy & Simon Rutt
SAP VERSION	1
SAP VERSION DATE	11/12/2017
EEF DATE OF APPROVAL	
DEVELOPER DATE OF APPROVAL	

Protocol and SAP changes

No changes since updated protocol was published.

1 'Catch Up[®] is the working name of The Caxton Trust, a not-for-profit charity registered in England and Wales (1072425) and Scotland (SC047557) as well as a company limited by guarantee (03476510). Catch Up is a registered trademark.'

Table of contents

Introduction	3
Randomisation	6
Outcome measures	7
Analysis	8
Report tables	11
References	12

Introduction

Catch Up Literacy is a structured one-to-one intervention for learners from the age of six to thirteen who are struggling to learn to read. The intervention involves two individual 15-minute sessions per week delivered by Teaching Assistants (TAs), teachers or mentors. The intervention adopts a combination of segmenting, blending phonemes and memorising letter names of high frequency sight words. It is targeted to the needs of individual learners, identified through a bank of formative assessments.

A member of staff within each school manages Catch Up Literacy while the intervention is usually delivered by trained TAs (Catch Up, 2015). Catch Up provides a four-part training programme for managers and classroom assistants, alongside ongoing support from Catch Up. It is intended that managers attend all four parts of the training, while classroom assistants (i.e. TAs) attend two.

There are four stages of Catch Up Literacy: assessments for learning, which are used to set targets and identify the appropriate starting points for pupils; selecting an appropriate book for the learner to read; delivery of two individual (one-to-one) sessions per week, each lasting 15 minutes; and ongoing monitoring, through which assessments for learning are revisited and targets are reviewed.

Catch Up Literacy was launched in 1998, based on original research by Diana Bentley, Suzi Clipson-Boyles and Dee Reid. It was designed for seven- to eight-year-old (Year 3) pupils who only achieved level 1 for reading in the Key Stage 1 Standard Assessment Tests, but has since been developed for use in secondary schools (Catch Up, 2008). It is now appropriate for learners from 6 to 14 years of age who struggle with reading.

Background evidence and significance

A previous efficacy trial funded by EEF evaluated the Catch Up Literacy intervention effect over control conditions. This trial focused on support that was delivered over the transition period between Year 6 and Year 7, with TAs delivering Catch Up to children at the end of Year 6 and up to two terms of Year 7. Outcomes from this trial suggested that, on average, pupils who received the intervention improved their literacy outcome (which was measured using the New Group Reading Test) by two months compared with the control group pupils. However, this difference was not statistically significant so we cannot be confident that it was not due to chance.

The purpose of this trial is to evaluate the effect of Catch Up Literacy over 'Business-as-Usual' (BaU) control schools in Key Stage 2. We propose to recruit 150 schools with, on average, eight eligible pupils from Year 4 and Year 5. Half of these schools will be randomly allocated to either the intervention or the BaU control group, making it a cluster randomised controlled trial. The primary outcome measure will be pupil's reading ability as measured by the [Hodder Group Reading Test](#) (HGRT) and the secondary outcome measure will use the [Salford Sentence Reading Test](#) (SSRT). Analysis will also explore, via pupil surveys, the effect of the intervention on pupil attitudes to school, and their confidence in and enjoyment of literacy.

The primary research question is:

What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' reading comprehension skills as measured by the Hodder Group Reading Test (HGRT)?

The secondary research questions are:

1. What is the impact of Catch Up Literacy on Year 4 and Year 5 FSM (those eligible to receive free school meals) pupils' reading comprehension skills as measured by HGRT?
2. What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' reading skills as measured by the Salford Sentence Reading Test (SSRT)?
3. What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' attitudes towards literacy?
4. What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' attitudes towards school?
5. What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' self-esteem?

Study design

Description of trial design

This is a cluster randomised controlled trial involving approximately 1200 pupils from 150 Primary schools. Schools selected up to 12 eligible pupils (the smallest cluster size is three pupils in one of the schools) from Year 4 and Year 5 and nominated two TAs to deliver the intervention (if they were randomised to the intervention group). Each school also identified a teacher to act as the project coordinator. If the school was randomised to the intervention group, three individuals (including the TAs and the coordinator) were offered the Catch Up Literacy training. The TAs from the intervention group delivered the programme to nominated pupils (between September/October 2016 and June 2017). As part of the intervention, TAs also assess these pupils at the end of each academic term to determine whether they are at their age expected levels in reading (but these assessments are not included in the independent evaluation). Once the pupils reach this level, they would stop receiving Catch Up Literacy. If the school was randomised to the BaU control group, they were asked to continue their teaching practices as normal. Following end-point testing in summer 2017 these control group schools will receive an incentive worth £790 which can either be in form of a grant or be used towards purchasing a programme of their choice.

Description of population

Catch Up was responsible for school recruitment. Between May and September 2016, Catch Up recruited 156 Primary schools from the North East region, Brighton, Cumbria, Grimsby, Hull & Immingham, Bournemouth and Plymouth. After expressing an initial interest to take part in the trial, schools were sent a memorandum of understanding (MoU) detailing roles and responsibilities of all parties involved and a formal consent of the headteacher was sought to participate in the trial.

Eligibility criteria

Primary schools that weren't already running Catch Up Literacy or Catch Up Numeracy were eligible to take part in the trial. Schools were asked to nominate up to 12 pupils whose reading age was below their chronological age.

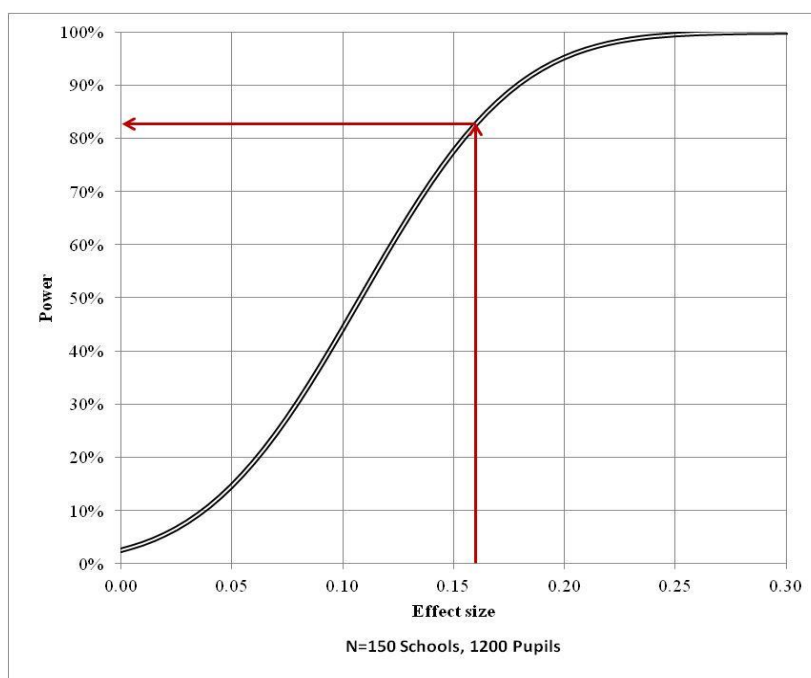
Once schools agreed to take part by signing a memorandum of understanding they were asked to send school and pupil data. This included school level information such as names and contact information for an individual to act as a coordinator and nominated TAs. Pupil data

included names, date of birth and unique pupil number (UPN) (administrative pupil data). As per the original design, baseline SSRT administration was to take place only after receiving administrative pupil data. However, due to the delays in most trial schools providing this data, baseline administration of SSRT had to take place simultaneously with the pupil data collection. This resulted in three schools where SSRT was administered but no administrative pupil data was provided. These schools did not receive their randomisation results.

Sample size

Randomisation was conducted at the school level. The intended sample size was 150 schools with an average of eight pupils each which is sufficient to detect an effect size of 0.16¹. Effect size of 0.16 was considered because this is the minimum effect size for which the approach is cost-effective (assuming a maximum of £80/pupil for 0.1 SD change and an approximate cost of £130 per pupil based on the previous trial) (Rutt, 2015). This minimum detectable effect size (MDES) is achieved at more than 0.8 power² by using the following assumptions: intra-cluster correlation of 0.137³; correlation between Key Stage 1 and Key Stage 2 of 0.73⁴ and average cohort size of eight pupils per school. This is also illustrated in Figure 1.

Figure 1: Power curve for a cluster design with 150 Primary schools



This design also allows for a small amount of attrition. If eight schools withdrew from the trial, we would still have a MDES of 0.16 with minimum statistical power of 0.8. We also expect that there will be schools that wouldn't put forward as many as eight pupils, which will be balanced by other schools who will put forward more than eight pupils to take part in the trial. Although we would expect that there will be approximately 170⁵ pupils in the trial who are eligible for free school meals, the size of the sample is not powered to run a separate FSM analysis.

¹ Bespoke MS excel was used to calculate the required sample size.

² Note that the effect size of 0.16 is achieved at 0.83 power. With the power of 0.8, the MDES reduces to 0.155.

³ This ICC is calculated using data from NPD 2013-2014 for key stage 2 reading total score and is referenced in EEF guidelines 2015 (EEF, 2015)

⁴ As referred in EEF's pre-test paper, the correlation coefficient of 0.73 is achieved by using KS1 as a covariate for a KS2 outcome measure (EEF, 2013)

⁵ This is based on the 14.1% of primary age pupils eligible for free school meals

Please note that we do not know the number of FSM pupils in the trial (as we are waiting for the national pupil database). Therefore, no prospective MDES is estimated.

Description of trial arms

Intervention group:

Catch Up Literacy is a structured one-to-one intervention for learners from the age of six to fourteen who are struggling to learn to read. The intervention involves two individual 15-minute sessions per week delivered by teaching assistants, teachers or mentors. The intervention adopts a combination of segmenting, blending phonemes and memorising letter names of high frequency sight words. It is targeted to the needs of individual learners, identified through a bank of formative assessments. A member of staff within each intervention school will manage Catch Up Literacy while the intervention will be delivered by trained teaching assistants. Catch Up provides a four-part training programme for managers and classroom assistants, alongside ongoing support from Catch Up. It is intended that managers attend all four parts of the training, while classroom assistants attend two. There are four stages of Catch Up Literacy: assessments for learning, which are used to set targets and identify the appropriate starting points for pupils; selecting an appropriate book for the learner to read; delivery of two individual (one-to-one) sessions per week, each lasting 15 minutes; and ongoing monitoring, through which assessments for learning are revisited and targets are reviewed.

Business-as-usual (BaU) control group:

Schools allocated to BaU control group will not receive any training and will not deliver Catch Up Literacy during the trial. Following end-point testing in summer 2017, these BaU control group schools will receive £790 which can be used either as a grant or towards purchasing a programme of their choice.

Number and timing of measurement points

The paper-based HGRT (primary outcome measure) will be administered and managed by NFER. Test administrators from NFER will be used to ensure that all HGRT tests are administered blind to treatment allocation and the same way in all schools, this will only take place at end-point in June-July 2017. A pupil attitude survey will be administered at the same time, after completion of the HGRT. It will be followed by administration of the SSRT (secondary outcome measure). This will be administered by research Assistants from University of Oxford at baseline (September- October 2016) and end-point (June-July 2017).

Randomisation

Two blocks of school randomisation were planned for this trial. It was anticipated that part of the sample will be recruited by the end of summer term 2016 with the remaining schools being recruited at the beginning of September 2016, when the academic year 2016-2017 started. The randomisation was planned to be a stratified randomisation using six strata, one for each of the five coastal areas and one for the North-East region altogether. However, due to the delays in receiving pupil data from the schools, the first block randomisation did not take place until September 2016, followed by two more blocks of randomisation, one in late September and the last in early October 2016. In order to reach the recruitment target, primary schools from Southend were also recruited to take part in the trial (this area was not considered originally in the protocol).

An NFER statistician carried out the stratified randomisation using geographical area as strata. There were seven strata (North-East region, Brighton, Barrow-in-Furness, Grimsby, Hull & Immingham, Bournemouth, Plymouth and Southend). Three blocks of randomisation were carried out by a statistician at NFER using a full SPSS syntax audit trail in September- October 2016. Table 1 identifies the final randomisation result by geographical area.

Table 1: Catch Up Literacy: results of the randomisation

Blocks	Strata	Randomisation group		Total
		Intervention	Control	
1	Cumbria	12	13	25
	Grimsby, Hull and Immingham	9	8	17
2	Brighton and Hove	4	4	8
	North East	37	36	73
	Plymouth	7	7	14
3	Bournemouth	6	7	13
	Southend	3	3	6
Total		78	78	156

As mentioned earlier, randomisation, pupil data collection and baseline SSRT administration took place simultaneously. This arrangement meant that NFER provided the randomisation results to Catch Up with a view to informing schools only when schools had submitted their pupil data. This resulted in three schools who were not informed of their group allocation (of these, two were intervention schools and one was a control school⁶). One further school was randomised due to an administrative error - this school never intended to take part and therefore was removed from the subsequent data collection. The resultant sample was 152 schools- 75 intervention schools and 77 control schools.

NFER organised the HGRT test administration in the summer term 2017. NFER test administrators (who were blind to group allocation) administered the tests in 71 intervention schools and 77 control schools. Four treatment group schools did not administer the HGRT test. These will be identified in the flow chart within the main report.

Outcome measures

Primary outcome

The primary research question is:

What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' reading comprehension skills as measured by the Hodder Group Reading Test (HGRT)?

The HGRT has been selected as the primary outcome measure which assesses pupils' reading comprehension at word, sentence and text levels. NFER's assessment experts felt it

⁶ Please note that two of these schools agreed to take part in the end-point testing and although they will be excluded from the main analysis, we will include them in the sensitivity analysis for the primary outcome measure.

was imperative that the measure used is primarily one of comprehension, as this is the point of reading. Decoding skills (in this case of single words) are clearly necessary but not sufficient in the development of comprehension. Unless the child is comprehending what s/he is reading, they will not succeed in the school system. It was also important that the measure linked well with KS2. For these reasons the HGRT was decided as the primary outcome. For this trial, raw total score (possible score range 1-53) from the HGRT II will be used, this test is designed for 7-12 years old pupils.

Secondary outcomes

The secondary research questions are:

1. What is the impact of Catch Up Literacy on Year 4 and Year 5 FSM (those eligible to receive free school meals) pupils' reading comprehension skills as measured by HGRT?
2. What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' reading skills as measured by the Salford Sentence Reading Test (SSRT)?
3. What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' attitudes towards literacy?
4. What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' attitudes towards school?
5. What is the impact of Catch Up Literacy on Year 4 and Year 5 pupils' self-esteem?

A separate FSM only analysis is planned for this trial to investigate the impact of the intervention on pupils receiving FSM (outcome measure 1 above). As noted above, proportion of FSM pupils is not yet available and therefore no prospective MDES is estimated.

Along with the HGRT assessment, pupils are also asked to complete a short survey to assess their self-esteem, attitudes to literacy and school. NFER aims to follow methodology adopted during the previous evaluation (Rutt, 2015) and create composite outcome measures listed in 3 - 5 above.

Once the primary outcome tests are completed pupils will sit the Salford Sentence Reading Test (SSRT) as a secondary outcome measure. This outcome was required as Catch Up felt a 1-to-1 administered assessment was a more reliable format to assess reading accuracy and comprehension. These tests will be delivered by Catch Up using 12 - 15 Research Assistants (blind to knowledge of group allocation) from University of Oxford and marked by Dr Ann Dowker of the same university (blind to knowledge of group allocation), before providing the dataset to NFER for analysis. NFER will use the raw scores from the tests, a reading score and a comprehension score, to analyse the outcome measure listed in 2 above.

Analysis

The trial analysis will follow EEF Analysis Policy⁷.

Primary intention-to-treat (ITT) analysis

The primary outcome analysis will be 'intention-to-treat'. Multilevel models with two levels (school and pupil) will be used for the analysis to account for the cluster randomisation. All schools with pupil data on the primary outcome measure will be included in this analysis irrespective to whether or not the schools implemented the intervention. The analyses will

⁷https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analyses_for_EEF_evaluations.pdf

determine whether the Catch Up Literacy intervention had an overall effect on pupils' reading comprehension skills.

The dependent variable for this model will be the raw total score on HGRT II with the following covariates:

- an indicator of whether the pupil is in the intervention school
- pupil prior attainment as measured by KS1 Reading attainment point score (KS1_READPOINTS variable on NPD 2013-14 (for year 5 cohort) and 2014-15 (for year 4 cohort))
- school's geographical location (representing the stratification variable used at randomisation).

In addition to the above models, we will also report a point estimate (without a confidence interval) from a similar model that doesn't include the stratification variable. This will be reported for the purposes of cross-study comparisons.

Imbalance at baseline for analysed groups

We expect no systematic bias to have arisen from randomisation. Furthermore, in cases where we do not have pupil administrative data (and therefore no NPD for such cases), the school doesn't know their group allocation and therefore can be considered unbiased dropout. We will obtain NPD data for all pupils for whom we have the administrative data and some of these pupils will be from schools that withdrew from the primary outcome measure. For these cases, we will be able to examine imbalance in the samples using pupil background characteristics such as pupil FSM status and prior attainment at Key Stage 1. The baseline differences will be presented as effect size as per the EEF guidance.

We will also look at school characteristics such as school attainment, school FSM, region and school type.

Missing data

We will run a multilevel logistic model with two levels (school and pupil) on whether or not a pupil is missing at follow-up, regressed on the covariates of the main model. As we are unlikely to be in a situation where a school has complete follow-up data and missing baseline, multiple imputation may not be useful. Instead, under the 'missing at random' given baseline assumption, we would expect a completer's analysis to be unbiased. If the school dropout is found to be larger than 5% from either of the two groups, we will conduct sensitivity analyses. This will be done by initially running multilevel multiple imputation. Following analyses undertaken on other EEF funded evaluations we would propose a methodology that includes all the variables included in the primary analysis plus other variables available from the NPD to run models that identify the significant variables associated with missingness. These significant variables would then be used for a multiple imputation process using the R package. The number of datasets is dependent on the amount of missing data but a minimum would be five, with a minimum of ten iterations. The model would then be extended using a weighting approach according to Carpenter et al. (2007). Missing data analysis will only be possible in cases where we have pupil administrative data and a subsequent match with the NPD.

Effects in the presence of Non-Compliance

The developer collected data on the level of school engagement throughout the delivery period via 'session logs' submitted by each TA. The logs include information such as the number of sessions each pupil has had, information on whether they reached the expected level and stopped receiving the sessions.

It is not possible to use the dosage data to identify compliance due to the 'stopping' strategy within the Catch Up treatment. Therefore, exploratory analysis will be undertaken using multi-level modelling techniques to identify the association between the amount of treatment received by an intervention group pupil and the outcome score. The outcome measure to be used will be raw score from the HGRT assessment. Independent variables to be included will be a measure of prior attainment using the raw score from the pupils' Key Stage 1 reading assessment, the schools geographical location (to account for the stratification carried out within the randomisation process) and instead of a flag to identify treatment assignment two additional variables will be entered that identify the number of treatment sessions a student received and whether the student stopped the intervention as they had reached the expected level of performance. The intervention guidelines have a clear protocol for this stopping process and all TAs receive guidance on this procedure at the formal training sessions. This additional term will be introduced as an interaction term within the model.

The model, whilst controlling for the effects of geographical region and pupil prior attainment, will be able to identify whether those pupils who reach expected levels before the end of the intervention delivery period, achieve, on average, higher HGRT scores than those pupils who are not part of the stopping strategy. The number of sessions and the interaction term will identify the association between the number of sessions received and the HGRT test outcome. With the interaction term included the variable for the number of sessions received identifies the impact of the sessions just for those students that didn't stop. This will remove the influence of the potential higher outcomes for those students that stopped receiving the intervention because they had reached the expected level of performance.

The exploratory procedure outlined above for the primary outcome will also be implemented for the secondary outcome measures.

Secondary outcome analyses

Completer's analysis will be run using secondary outcome measures as dependant variables in two multilevel models. The dependent variables for these models will be the raw reading and comprehensions scores from SSRT⁸ with the following covariates:

- an indicator of whether the pupil is in the intervention school
- prior attainment as measured by baseline SSRT raw scores for reading and comprehension
- school's geographical location (representing the stratification variable used at randomisation).

⁸ As mentioned previously, SSRT was required as Catch Up felt a 1-to-1 administered assessment was a more reliable format to assess reading accuracy and comprehension for struggling readers. Baseline SSRT scores will be used instead of KS1 as a prior attainment measure for this outcome. Although baseline SSRT was administered at different time than the KS1 (but prior to knowledge of group allocation), we do not envisage these will be systematically different for the randomised groups.

Similar multilevel models will also be run with pupil attitude measures as dependant variables where the covariate will be the KS1 Reading attainment point score (KS1_READPOINTS variable on NPD) instead of the baseline SSRT score to be consistent with the specification used for the primary model. If the secondary outcome measures encounter attrition, it will be important to determine the extent of bias. Unlike the primary outcome measure where the covariate is KS1 and unlikely to be missing with complete follow-up, the covariate in the SSRT model is the baseline SSRT which might be affected by missing data. If there is more than 5% data missing at baseline where we have the outcome measures at follow-up, multilevel multiple imputation will be used to impute the missing values at baseline. The imputation will follow the same methodology identified in the section on missing data. This model will be compared with the completer's model.

Data manipulation will be carried out in SPSS while the multilevel models will be run in MLwiN using imputation macros available from missingdata.org.uk.

Subgroup analyses

Sub-group analyses on the primary outcomes will be carried out as per the protocol and the most recent EEF analysis guidelines. A separate analysis of FSM only pupils will be carried out as per the EEF analysis guidance. These models will be similar to the main models of overall effect but will only include pupils who were eligible for FSM as measured by EVERFSM_6 variable. This FSM variable is the preferred option as it identifies pupils who may not be eligible for free school meals now, but who have been in the recent past, normally the previous six years. This is because the effect of FSM eligibility can have longer lasting effects even after a pupil is no longer eligible.

As per the protocol, we will explore the differential effect based on pupil age (as a continuous variable), gender and pupil FSM status (whether a pupil has ever received free school meals as measured by EVERFSM_6 variable) in three separate interaction models. As per the EEF guidance, this will be done using models identical to the main models but including the interaction and the intervention indicator as covariates.

Effect size calculation

The numerator for the effect size calculation will be the coefficient of the intervention group from the multilevel model. All effect sizes will be calculated using total variance from a multilevel model, without covariates, as the denominator i.e. equivalent to Hedges' g.

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

Confidence intervals for each effect size will be derived by multiplying the standard error of the intervention group model coefficient by 1.96. These will be converted to effect size confidence intervals using the same formula as the effect size itself.

Report tables

All the tables will be structured according to the EEF trial report template⁹.

⁹ <https://educationendowmentfoundation.org.uk/evaluation/resources-centre/writing-a-research-report/>

References

Angrist, J.D. and 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. [online]. Available: <http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476535> [19 May, 2017].

Carpenter, J.R., Kenward, M.G. and White, I.R. (2007) Sensitivity analysis after multiple imputation under missing at random – a weighting approach. *Statistical Methods in Medical Research* 16:259-275. Dunn, G, Maracy, M and Tomenson, B (2005). Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Statistical Methods in Medical Research* 2005; 14: 369-395

Education Endowment Foundation (2013). *Pre-testing in EEF Evaluations*. London: EEF [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/Pre-testing_paper.pdf [3 August 2017]

Education Endowment Foundation (2015). *Intra-cluster Correlation Coefficients*. London: Education Endowment Foundation [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_ICC_using_NPD_data.pdf [3 August 2017]

Rutt, S., Kettlewell, K. and Bernardinelli, D. (2015) *Catch Up[®] Literacy: Evaluation Report and Executive Summary*. NFER: Slough. [online]. Available: <https://www.nfer.ac.uk/publications/EFCL01/EFCL01.pdf> [01 June, 2017].