Statistical Analysis Plan for Achievement for all
University of Manchester

Education Endowment Foundation

INTERVENTION	Achievement for All Schools Programme
DEVELOPER	Achievement for All 3As
EVALUATOR	University of Manchester
TRIAL REGISTRATION NUMBER	DOI 10.1186/ISRCTN67347514
TRIAL STATISTICIAN	Dr Sophina Choudry and Dr Elizabeth Byrne
TRIAL CHIEF INVESTIGATOR	Dr Garry Squires and Prof Neil Humphrey
SAP AUTHOR	Dr Sophina Choudry and Dr Elizabeth Byrne
SAP VERSION	Version 3.1
SAP VERSION DATE	03/04/2019

Protocol and SAP changes

Primary intention-to-treat (ITT) analysis

We will apply a two-level (school, pupil) multi-level model rather than a three-level multi-level model with time as the third level (as set out in original the protocol). Including time as a third level would require more than one measurement point for the analysis to be feasible, and so it has been excluded from the analysis, along with any cross-level interaction term (group*time; e.g. if AfA, if post-test). The type of multi-level model being fitted has also been amended from a random slopes and random intercepts model, to a fixed effects and random intercepts model. Our reason for making this change is that we have no prior theoretical expectation that there will be differences in the slopes of different schools. A fixed effects model is suitable for answering the research questions. We also include 'AfA target group' in the subgroup analysis. We have done so because the AfA program has a specific focus on raising attainment for this particular group of students. We have also changed the primary outcome from KS2 total marks (for reading and writing) to KS2 reading marks only. This change has been requested by EEF following the introduction of new SATs and questions about the reliability of the writing component. Finally, we have changed the software package being used to run the multi-level analysis from MLWin to Mplus.

Sensitivity analysis

Model 1.1 is the main analysis from which we will draw our primary conclusion about the effectiveness of AfA, and 1.2 onwards are part of the sensitivity analyses. The randomisation variables have been removed from the main ITT analysis (model 1.1) and added to the sensitivity analysis (model 1.2) to ensure our analyses align with EEF's requirements and to justify the inclusion of covariates statistically. The statistical method being used to account for missing data has been changed from multiple imputation (MI) to full information maximum likelihood (FIML).

Please note that we have now also included model 1.3 and model 1.4, which have KS2 total marks and KS2 writing as outcome variables respectively. We will also check for correlation between reading and writing in order to address the concerns raised regarding low correlation in recent years (see Allen, 2016).

Statistical Analysis Plan for Achievement for all
University of Manchester

Table of contents

Introduction

Intervention

The Achievement for All (AfA) Schools Programme is a national school-based programme aiming to address the gap in attainment between the lowest achieving 20% of children and their classmates. It is essentially a school improvement programme that focuses on four areas: leadership and governance in schools; teaching and learning; parent and carer engagement; and, wider outcomes and opportunities.

The direct recipients of the AfA Schools Programme are the teachers in a school. Each AfA school has a designated member of staff known as the AfA champion who works in collaboration with an AfA coach (employed by AfA 3As) to identify the needs of the school in order to develop a tailored program. Together they coordinate and select priority school/teacher development areas out of the four mentioned above. Each AfA school also has a lead teacher who may be the same person as the AfA champion (but not necessarily). A step-by-step guide is subsequently provided for participating schools by the AfA coach. In addition, teachers are able to access an online learning platform (The Bubble) containing four core, tailored (i.e. modules that have been selected by AfA coach based on the school's priority development area from a list of options – e.g. leadership for inclusion → collaborative action research, etc.) and partner modules (e.g. BBC Children in Need fun and friendship, etc.). The four core modules are Leadership for Inclusion; Teaching and Learning; Wider Outcomes and Opportunities and Engaging with Parents and Carers. Each of these core development areas contains a core module and a list of 'tailored' modules. For example, the core area Engaging with Parents and Carers has one core module, namely, 'structured conversations with parents', and two tailored modules called 'early support' and 'welcoming and including families'. Teachers have the opportunity to access these modules online to negotiate their own learning or have the AfA coaches deliver them. Each AfA school gets up to 12 AfA coach 'interactions' in each school year. The designated AfA coach supports teachers in their attempts to enact changes in practice in the above areas. Schools taking part in the intervention also have the opportunity to work towards an AfA quality mark based on an evaluation by the AfA coach of progress made, measured against a set of AfA derived standards. Not every school taking part in the AfA programme is awarded the quality mark.

Trial

This effectiveness trial comprises two distinct strands: (i) the randomised controlled trial (RCT) strand, and (ii) the implementation and process evaluation (IPE) strand. The evaluation team are responsible for (a) initial consultation and guidance to ensure suitable groundwork for a robust trial (i.e. advising how many schools AfA may wish to recruit, ethical considerations, research design); (b) working with identified schools to participate in data collection and analysis as part of the research for both the RCT and IPE strand; and (c) analysis and report writing. Achievement for All (AfA 3As) are responsible for recruitment of schools to a pool of potential participants and for delivery of the AfA Schools Programme.

Purpose

In the national pilot, the AfA Schools Programme was shown to be very promising, particularly in relation to vulnerable learners in the education system (including those with SEN and those from disadvantaged backgrounds). However, there is genuine uncertainty about the strength of the claims made around the potential of the programme to produce socially significant change in attainment and other outcomes for children

The primary aim of the RCT strand of the study is to explore whether tailored whole school approaches delivered by Achievement for All (AfA 3As) can lead to improvements in children's academic performance in literacy (specifically, those in Years 4 and 5 2016/2017). The secondary aim of the RCT strand will be to explore whether there are improvements in maths, attendance, and resilience-related outcomes.

The aim of the IPE strand of the study is to understand and unpack the complex relations that underpin such a flexible, multifaceted model of implementation, where the tailored packages delivered by Achievement for All (AfA 3As) for one school are likely to be different to the package for another school.

We will try and understand the process of delivery of the intervention through a case study approach, where schools will be the cases. This will allow us to account for exemplary practices of the model delivered, how these vary from case to case, and with what consequences for students' attainment and other outcomes. If there is no overall impact of using Achievement for All (AfA 3As) compared to usual school improvement practices, then the IPE strand will help us to understand why this might be.

Research Questions

This RCT strand of the trial is designed to answer the following research questions:

1. Compared to usual practice, what is the impact of AfA on children's literacy (primary outcome), maths, attendance[1], and resilience-related outcomes (secondary outcomes)?
   a. After 5 terms of exposure (Year 5 cohort) (quant)
   b. After 6+ terms of exposure? (Year 4 cohort) (quant)
   c. What are the perceived impacts of AfA among intervention stakeholders (e.g. teachers, head teachers)? (qual)

2. In relation to RQ1 above, are there differential intervention benefits in the above outcomes among pre-specified subgroups of children? (quant/qual)
   a. Among children eligible for free school meals (FSM)? (quant)
   b. Among the target group of children identified by participating schools as belonging to 'the lowest achieving 20%'? (quant)
   c. What processes underpin any differential intervention benefits identified? (qual)

The IPE strand of the trial is designed to answer the following research questions:

3. How is AfA implemented, and what difference does it make? (quant/qual)
   a. How and why does AfA implementation vary? (quant/qual)
   b. To what extent does implementation variability moderate intervention outcomes?
      i. Do outcomes vary as a function of 'on treatment' status? (quant)
      ii. Do differential intervention benefits among specified subgroups vary as a function of 'on treatment' status? (quant)
      iii. What are the proposed critical components of AfA, and to what extent does their relative presence/absence influence outcomes? (quant/qual)
   c. To what extent does contextual variation influence the implementation of AfA (and, subsequently, outcomes)? (quant/qual)
      i. How and why is this the case? (qual)
4. Is there evidence to support the AfA theory of change? (quant/qual)

Please note that RQ1c and RQ2c will be dealt with qualitative analysis of case study data obtained during the IPE phase.

Study design

Population and Eligibility Criteria

The plan in the protocol was to recruit 160 primary schools by AfA. In the first instance AfA attempted to recruit schools in the North East of England and then broadened out to other regions as required. AfA's planned recruitment strategy included 'talking head' videos, presentations, and network events. In terms of eligibility, schools already (or previously) involved in the AfA programme (or its pilot) were excluded. Overall, the expectation was that the school sample would mirror the national EverFSM average for Key Stage 2 (c.30%).

The target figure of 160 schools exceeded the 140 needed for the trial to be adequately powered (see Power Analysis and Sample Size [PASS] calculations below) by allowing for some attrition at the baseline data collection stage.

---

[1] Among children in the AfA target group ('the lowest achieving 20%').

Within the schools, our target cohorts were children beginning Years 4 (RQ1b) and 5 (RQ1a) in September 2016. Assuming a cluster size of approximately 40 (based on two previous trials led by the evaluation team and also AfA's records from their current primary school customers), the total sample size in each cohort was expected to be N=4,800, of which we anticipated subgroups of between 1,440 (RQ2a – 30% of sample expected to be eligible for FSM) and 960 (RQ2b – '20%' target group). The exact composition of the 20% target group is discussed and agreed with each school during the initial needs analysis conducted at the beginning of the intervention. However, as this target group needed to be identified in *all* schools *prior* to randomisation, participating schools in this trial nominated their 20% target group as part of the process of signing up for the evaluation using guidance provided by AfA.

Sample Size

At the randomisation stage, 134 schools were recruited to the trial that met the eligibility criteria as per the Memorandum of Agreement. Despite not achieving the target or recruiting 140 schools, the trial is adequately powered. There are a total of 12924 pupils (Year 5 n = 6338; Year 4 n = 6586) in the study. From the data it is evident that 1374 Year 5 and 1350 Year 4 pupils have been nominated as the target cohort (22% and 21%, respectively). The main intention to treat (ITT) analysis and the subsequent EEF report will be written based on the Year 5 data set. The Year 4 data analysis serves the purpose of validating the main findings and will be treated as an addendum to the main EEF report. For that reason, this SAP is written for the Year 5 data primarily but also applies to the Year 4 data.

Trial Design and Trial Arms

The trial design is a two-arm cluster randomised controlled trial (RCT) incorporating a comprehensive implementation and process evaluation (IPE). There are two arms: intervention (AfA School Programme) and control (business as usual). Schools are the unit of randomisation and have been assigned to implement AfA or continue usual practice throughout the evaluation period. Those schools who are assigned to the usual practice arm will receive a retention incentive of £1000 (to be paid in instalments as follows: £200 following random allocation, £200 at the end of the first year of the trial, £200 at the midpoint of year 2 of the trial, and £400 at the conclusion of the trial and on completion of required data/surveys) in order to minimise differential attrition.

For the IPE strand has three phases:

- Phase 1 – Case Study Selection Process: We carried out a quantitative analysis to identify and select the case study schools using the baseline data obtained as part of the RCT strand from the schools (see section 'additional analysis' for more detail).
- Phase 2 – Longitudinal Case Studies: This phase involves a longitudinal case study analysis of the 8 schools identified through Phase 1 above - planned for the duration of the project (two school years). The unit of analysis of the observations is the teacher's practice. That is, we seek to understand the extent to which the teachers begin to apply the principles of the AfA Schools programme in their practice (including planning of lessons and teaching them - see protocol for further details).
- Phase 3 – School Implementation Survey: This phase will be administered towards the end of the second year of the project (June/July, 2018) and involves the administrating an implementation survey to each of the 66 intervention schools. The surveys themselves will be developed during the first year of the project based on the qualitative data collected in Phase 2 (see above).

Number and Timing of Measurement Points

Table 1 – Timeline for trial data points (RCT and IPE)

Activities	Lead-in 03/2016-05/2016	09/2016	10/2016	05/2018	06/2018	07/2018	08/2018	05/2019
Primary outcome measures (Y4 cohort)	KS1 pre-test (sat in summer 2015)							KS2 post-test
Secondary outcome measures (Y4 cohort)	% attendance 16/17	Resilience pre-test			Resilience post-test		% attendance 17/18	
Primary outcome measures (Y5 cohort)	KS1 pre-test (sat in summer 2014)			KS2 post-test				
Secondary outcome measures (Y5 cohort)	% attendance 16/17	Resilience pre-test			Resilience post-test		% attendance 17/18	
Usual Practice Survey (UPS)		School UPS			School UPS			
Implementation Survey					School IS			
Case Study Data collection		03/2017 – 06/2018 (for Year 5)						

Randomisation

Randomisation took place in November, 2016 after the baseline pre-test surveys for the secondary outcome of 'resilience' were completed. The randomisation procedure was conducted independently of the evaluation team by the Manchester Academic Health Science Centre Clinical Trial Units (MAHSC-CTU) to eliminate selection bias. A minimisation algorithm was applied, utilising the following school-level co-variates sourced from EDUBASE: %FSM, %SEN, and Attainment[2]. Given the nature of the AfA intervention and the primary trial outcome, these are the most important variables on which to obtain balance at baseline. As a result of the randomisation process, and in order to achieve balance based on the minimisation protocol, 66 and 68 schools were randomly allocated to the intervention arm and control arm respectively (rather than 67 in each group).

Calculation of sample size

PASS calculations are based on our primary outcome measure using the software Optimal Design Program (version 3.01) as recommended by EEF. At the protocol stage, Achievement for All (3As) aimed to recruit 140 schools. The following assumptions were taken into account in the power calculations (please note that these are MDES calculations for Year 5 data only, as the main report will be written based on this dataset):

---

[2] Any such information missing from EDUBASE will be sourced directly from the school in question.

Table 2 – MDES Estimates – Predicted Sample

Cohort	No. of clusters	Average cluster size	Pre-Post test correlation[3]/R^2	ICC[4]	Power	Significance	MDES
Whole Cohort	140	40	0.7/0.49	0.14	0.8	0.05	0.15
FSM Cohort	140	12	0.7/0.49	0.14	0.8	0.05	0.18
AfA Target Cohort	140	8	0.7/0.49	0.14	0.8	0.05	0.20

In the actual sample, a total of 6338 pupils over 134 schools were included. Thus, the power calculations for the randomised sample are as follows:

Table 3 – MDES Estimates – Achieved Sample

| Cohort	No. of clusters	Average cluster size (Year 5)	Pre-Post test correlation/R^2	ICC[5]	Power	Significance	MDES
Whole Cohort	134	47	0.7/0.49	0.07	0.8	0.05	0.12
FSM Cohort	134	14	0.7/0.49	0.07	0.8	0.05	0.18
AfA Target Cohort	134	10	0.7/0.49	0.07	0.8	0.05	0.20

---

[3] Assumed pre-post correlation of 0.7 (based on KS1-KS2 correlation in the NPD). Thus, R^2 will be 0.49.
[4] EEF analysis of KS2 data from NPD.
[5] Calculated based on KS1 literacy point scores obtained from the NPD data.

Follow-up

Figure 1: Participant flow diagram for Year 5
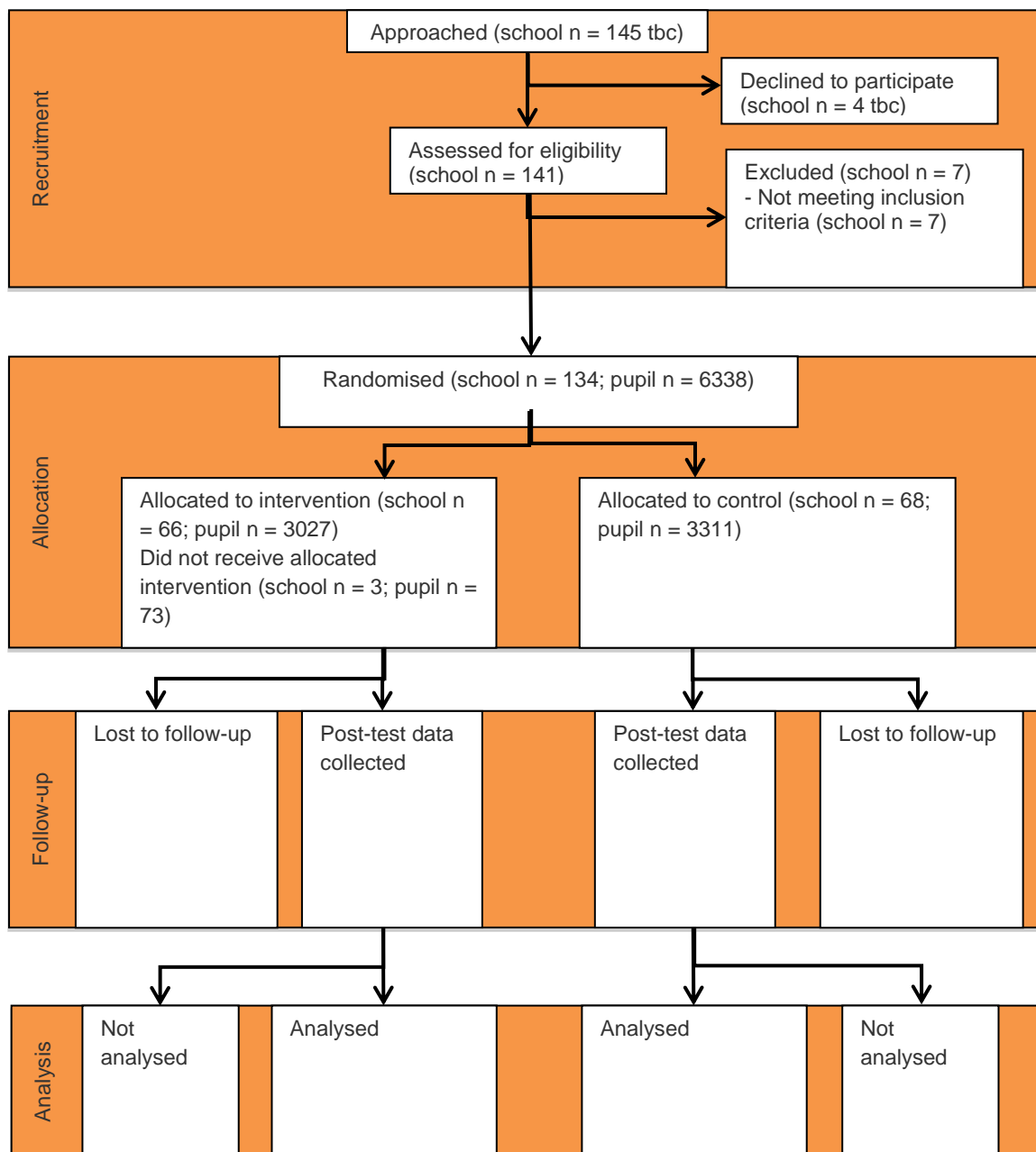
Outcome measures

Primary outcome

The primary outcome measure for the trial is children's academic attainment in reading for the Year 5 cohort. Specifically, we are interested in the impact of AfA on the reading levels of (a) the AfA 'target group' and (b) the wider population of children in our trial cohort as joint primary outcomes.

This data will be sourced from the National Pupil Database (NPD), with end of Key Stage 1 data used as a pre-test covariate and end of Key Stage 2 data used as the main post-test outcome. Using NPD data minimises attrition and bias, and in particular, preserves the ITT analyses without having to account

for missing data. This greatly reduces the data burden on participating schools, and increases the external validity of the evaluation (since this data provides the primary metrics by which schools in England are judged). The variables used are:

- ***Post-test - KS2 English Reading Marks (marks for reading only)***
- ***Pre-test - KS1 Literacy Point Score (reading and writing)***

Secondary outcomes

Attendance

Attendance data (% half-days missed due to unauthorised absence) will be sourced from the NPD. Given the uniformly high attendance rates across primary schools, our analyses of this outcome will be restricted to children in the AfA target cohort.

Resilience-related outcomes

Children's self-reported *self-esteem*, *goals and aspirations*, *family connection,* and *school connection* will be assessed using subscales of the Student Resilience Survey (SRS; Sun & Stewart, 2007). Pupils read ten statements (e.g. "I can do most things if I try") and respond on a 5-point scale (where 1 = Never and 5 = Always). These outcomes will be captured via a secure online survey platform (World App Key Survey). The domains or sub-scales to be assessed were agreed in discussion between the University of Manchester, AfA and EEF as being those that provided the optimal fit to non-academic outcomes noted in the AfA theory of change. The SRS is a relatively new instrument and although sub-scales have been identified by Sun and Stewart (2007), we will need to check the consistency of the scales within our data set. We will be carrying out a preliminary analysis using the baseline SRS data in order to establish the factor structure of the items and validate the sub-scales. On the basis of this analysis we may revise the subscale structure.

Other outcomes

Academic attainment in writing and maths for Year 5 pupils (and Year 4 reported in the appendix) will also be modelled.

Analysis

Primary intention-to-treat (ITT) analysis

RQ1: Compared to usual practice, what is the impact of AfA on children's reading (primary outcome), writing, maths, attendance[6] and resilience-related outcomes (secondary outcomes)?

> **a) After 5 terms of exposure (Year 5 cohort) (quant)**
> **b) After 6+ terms of exposure? (Year 4 cohort) (quant)**

To address RQ1 an ITT analysis will be conducted for the primary outcome measure (KS2 reading) using multi-level modelling (MLM) with fixed effects and random intercepts. To account for the nested nature of the data, hierarchical models with two levels (school, pupil) will be fitted (controlling for pre-intervention/KS1 scores at the pupil level). Mplus (Version 8.2) will be used to conduct the analysis.

At the school level we will only include group (e.g. AfA, usual practice) as an explanatory variable. At the child level we will include prior KS1 literacy attainment as an explanatory variable given its established association with our primary trial outcome. We will refer to this model as Model 1.1. If the co-efficient associated with the trial group variable is statically significant, an intervention effect will be noted. Subsequently, we will carry out sensitivity analysis by adding randomization variables, the usual

---

[6] Among children in the AfA target group ('the lowest achieving 20%').

practice indicator(s), Special Educational Needs (SEN), and gender to Model 1.1 (see sensitivity analysis section on page 11 for further details).

<div align="center">**Equation 1 – Primary Outcome Analysis**</div>

$$Y_{ij}^{Post} = \alpha + \beta.Group_j + \gamma.Y_{ij}^{Pre} + \theta.Rand_j + \delta.X_{ij} + \varepsilon_{ij}$$

Where:

Y^{Post} = KS2 reading scores (for the primary outcome analysis)
$\underline{Y^{Pre}}$ = Key Stage 1 literacy point scores
Group = Condition at school level *j*
Rand = Randomisation variables at the school level *j* (i.e. %FSM, %SEN, and %RWM 4+ categorised as low, middle, and high for the sensitivity analysis only)
X = control variables at the pupil level (e.g. gender, SEN for the sensitivity analysis only)
ε = error term for pupils *i* clustered at school level *j*
i = pupil *i*
j = school *j*

Interim analyses

N/A

Imbalance at baseline

As the analysis is based on ITT, even if schools allocated to the intervention arm do not receive the AfA schools programme (see Participant Flow Diagram above), their data will be included in the analysis. The primary outcome variable (KS2 literacy attainment) will be obtained from the NPD, and thus attrition will not affect balance at baseline. The 3 schools identified to date that are not receiving the intervention have agreed for the data to be collected, analysed and reported. We will demonstrate equivalence at baseline for the primary and secondary outcomes. This process has already started for data that has already been obtained (e.g. KS1 attainment data – see tables on page 14 and onwards of this SAP).

Missing data

The extent of missing data for the intervention and control arms, which includes observations missing due to school attrition and other reasons such as pupil absence on the day of the tests or incomplete tests, will be established once post-test data has been obtained. The % of incomplete cases for each outcome in the dataset will be determined. If this exceeds 5% we will perform additional sensitivity analyses by estimating the models using the Full Information Maximum Likelihood (FIML) method to deal with missing data.

Missing data will be further investigated by introducing a binary variable for complete (1) and incomplete cases (0). Here, a complete case is defined as a case that has the post-test (KS2 literacy score) and pre-test (KS1 literacy score) data present. We will carry out a regression analysis with this as an outcome variable and condition (treatment or control), FSM eligibility and AfA target cohort (Yes/No) as explanatory variables. By doing so, we will be able to deduce students (e.g. who are eligible for FSM) likelihood of completing the pre and post-test and thereby their likelihood of having complete cases.

Subgroup analyses

RQ2: In relation to RQ1 above, are there differential intervention benefits in the above outcomes among pre-specified subgroups of children? (quant/qual)
      a) Among children eligible for free school meals (FSM)? (quant)
      b) Among the target group of children identified by participating schools as belonging to 'the lowest achieving 20%'? (quant)

Subgroup analyses will be conducted to address RQ2. The model presented in equation 1 (Model 1.1) for the primary and secondary outcome measures will be re-run twice with a subset of the main data.

First, only FSM pupils will be entered, resulting in Model 2.1, and then only AfA target pupils will be entered, resulting in Model 2.2.

Sensitivity Analysis

The sensitivity analysis will be conducted in Mplus (Version 8.2) and will involve various steps. First, a number of explanatory variables will be added to Model 1.1 (resulting in Model 1.2 below). At the school level the randomisation variables (%FSM, %SEN, and %KS2 Reading Writing Mathematics combined level 4+; given that these were the minimisation variables, as per the EEF protocol) and the usual practice indicator(s) (derived from the Usual Practice Survey in order to provide a more robust estimate of the achieved relative strength of AfA) will be entered as co-variates. At the pupil level SEN and gender will be added as co-variates. If the co-efficient associated with the trial group variable is statically significant, an intervention effect will be noted. This will be converted to Hedge's *g* accounting for varying cluster sizes, as per EEF reporting guidelines. The change in the log-likelihood and changes in variance partition coefficient (VPC) between models will also be reported.

- Model 1.2 – will involve the same specification as Model 1.1 plus some additional school level (i.e. the usual practice indicator(s), the randomisation variables: %FSM, %SEN and %KS2 Reading Writing Mathematics combined level 4+), and pupil level (i.e. SEN, gender) co-variates.

Next, if the proportion of incomplete cases for the particular outcome variable being used in the model exceeds 5% (see the Missing data section above), FIML will be used to re-estimate models in MPlus (Version 8.2) to deal with missing data. Therefore, partially observed cases can be included in the sensitivity analysis, for example cases that either have a missing KS2 literacy test score or a missing KS1 literacy test score. This would involve repeating the previously described analyses (including the primary ITT and subgroup analyses, and the addition of the co-variates) using FIML to estimate the models.

The final step of the sensitivity analyses will involve repeating the whole analyses described above (Model 1.1 and 1.2, and Models 2.1 and 2.2) for two additional measures as the outcome variables. For model 1.1 (main ITT model), we have used KS2 English total marks for reading only as our primary outcome variable. This is because evidence has shown that recent changes to assessment have led to lower correlation between KS2 reading and KS2 writing (from 0.84 in 2015 to 0.34 in 2016; see Allen, 2016). Hence, we propose a sensitivity analysis in which we re-run the main ITT with KS2 English (reading and writing) and also KS2 writing alone:

- Model 1.3 – will include KS2 English total marks (reading and writing) as a secondary outcome variable.
- Model 1.4 – will include KS2 writing as a secondary outcome variable.

We will also check and analyse the attainment data that will be collected in 2018 and 2019 for a changes in the correlation between KS2 reading and writing.

Implementation analysis

RQ3: How is AfA implemented, and what difference does it make?
       a) How and why does AfA implementation vary?
       b) To what extent does implementation variability moderate intervention outcomes?
              i) Do outcomes vary as a function of 'on treatment' status?
              ii) Do differential intervention benefits among specified subgroups vary as a function of 'on treatment' status?
              iii) What are the proposed critical components of AfA, and to what extent does their relative presence/absence influence outcomes?

Using the knowledge generated from our IDEA workshop and the first three terms of our longitudinal implementation case studies, we will design a school-level implementation survey to be administered through a secure online portal during the second year of the trial (2017/18) and completed by the

school's AfA lead. This survey will be used to (a) document implementation activity across the schools in the intervention arm of the trial, thus enabling us to (b) assess the relative presence/absence of proposed critical components of AfA, and (c) document hypothesised change mechanisms/processes. The data from the implementation survey will be used to measure and construct implementation variables.

For RQ3a we will use data from schools allocated to the intervention arm of the trial. Two-level MLM will be employed (pupils clustered in schools). Only schools in the intervention arm will be included in this analysis (i.e. 66 schools). The basic model (model 3.1) will be fitted for the primary outcome measure (KS2 reading) and follow the same specification (including accompanying explanatory variables) as specified in equation 1 – primary outcome analysis. The only difference will be the removal the Group variable (if intervention or control). In addition, we will introduce implementation variables to model 3.1 to determine the effect of implementation on the outcome variable. To address RQ3b(i), on-treatment status will be introduced as a school level explanatory variable. We will collect data from AfA to ascertain how many treatment schools will have been awarded with their Quality Mark (QM) status[7] by the end of their program. Schools that are awarded a QM will be coded as *on-treatment*. Schools that will not have achieved QM will be coded as *off-treatment*. To address RQ3b(ii), we will follow the above procedure but in addition introduce the sub-group membership indicator (e.g. FSM eligibility) as an explanatory variable at the pupil level. This will allow us to model cross-level interaction (on-treatment*sub-group, e.g. if on-treatment, if FSM) to determine whether differential intervention benefits among specified subgroups vary as a function of on-treatment implementation. For RQ3b(iii) we will introduce critical component indicators as explanatory variables at the school level and model outcomes as specified above.

Changes in the log-likelihood and variance partition coefficient (VPC) will be reported to determine whether the introduction of these variables or the interaction explains any further variation in the model.

Structural Equation Modeling

RQ3c: To what extent does contextual variation influence the implementation of AfA (and, subsequently, outcomes)?

RQ4: Is there evidence to support the AfA theory of change? (quant/qual)

For RQ3c and RQ4 we will employ Structural Equation Modeling (SEM), specifically, multi-level path analysis using the software Mplus (version 8.2). We will analyse two relations in the SEM path analysis (see figure 2):

- The relation between contextual variations and the outcome;
- and if this relation is mediated by implementation variables.

The contextual variation variables are the school level randomisation parameters (%FSM, % SEN, %RWM 4+ categorized as low, medium and high). The outcome is KS2 reading scores and the implementation variables are yet to be constructed (see above).

Mplus Model Specification: Context implementation outcomes paths
Type: Two-level model clustered at the school level
Analysis: (if categorical variable) robust weighted least squares (WLSMV)

---

[7] The AfA schools programme primary school quality scheme as two awards: Quality Mark (QM) and Quality Lead (QL). All school who successfully complete the AfA programme are expected to achieve QM status. This is evaluated in the second year of the programme against AfA extended criteria. AfA's estimate is that c.50% of schools will achieve QM status within the time period of this trial.
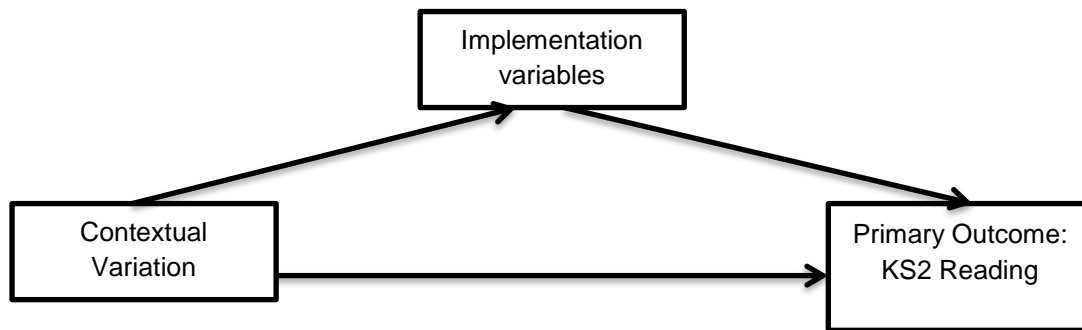
Figure 2 - Mplus Model Specification – relations explored

The findings from this analysis will then inform RQ3c, as well as RQ4 by comparing the result to the existing AfA theory of change.

Secondary outcome analyses

The secondary outcome analysis will follow the same procedure as demonstrated in the primary and sensitivity analyses (see equation 1 above), but with different dependent variables. Therefore, there will be different models for each of the following measures:

- Model 1.1.1: attendance with a binary variable of up to 10% being coded as 0 and more than 10% being coded as 1. Please note that the threshold for 10% is being applied as per the Department for Education definition of persistence absence (see DfE, 2017).
- Model 1.1.2: resilience related outcome, self-esteem (SRS).
- Model 1.1.3: resilience related outcome, goals and aspirations (SRS).
- Model 1.1.4: resilience related outcome, family connection (SRS).
- Model 1.1.5: resilience related outcome, school connection (SRS).
- Model 1.1.6: KS2 mathematics attainment with KS1 mathematics points as an individual background variable.

The primary and secondary outcome analysis will address RQ1.

Additional analyses

Besides basic descriptive analysis (e.g. mean, standard deviation for control and intervention groups), we also carried out a two-level multi-level model analysis using the baseline data obtained as part of the RCT strand from the schools (through AfA) with the aim to identify case study schools (phase 1 of IPE strand).

We applied a two-level multi-level model (random intercept and slope model), where pupils are clustered at school level:

Equation 2 – Case Study Selection Model

$$Y_{ij} = \alpha + \delta.X_{ij} + \varepsilon_{ij}$$

Where:

Y = Key Stage 1 literacy point scores (for the primary outcome)

X = control variables (i.e. Free School Meal eligibility, Special Educational Needs, and AfA target group)

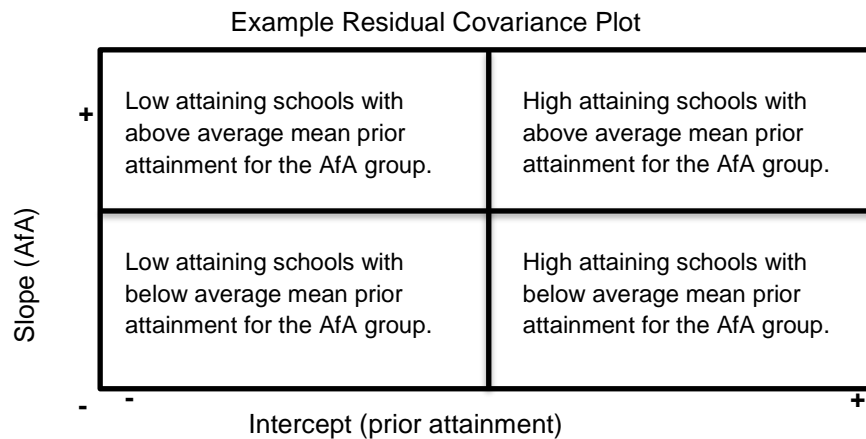*ε* = error term for pupils clustered at school level

i = pupil *i*

j = school *j*

All intervention schools (n = 66) in the subset database were ranked according to their residuals and relevant graphs were produced.

- The ranking took place according to the following criteria:
  i.   Low attaining - schools with greater proportion of students with below average KS1 Literacy scores.
  ii.   High attaining - schools with greater proportion of students with above average KS1 Literacy scores.
  iii. Adding more value to the AfA group - schools with greater proportion of AfA students with above average KS1 literacy point mean.
  iv.   Adding less value to the AfA group - schools with greater proportion of AfA students with below average KS1 literacy point mean.

- The residual analysis from the MLM will lead to the production of a residual covariance plot (see figure 1).

Figure 1 – Illustrative example of each quadrant in a residual covariance plot – normal MLM models

Example Residual Covariance Plot

	Low attaining schools with above average mean prior attainment for the AfA group.	High attaining schools with above average mean prior attainment for the AfA group.
	Low attaining schools with below average mean prior attainment for the AfA group.	High attaining schools with below average mean prior attainment for the AfA group.

Slope (AfA)

Intercept (prior attainment)

We identified 12 schools, 3 from each quadrant in figure 1, according to highest rank in the specific category and contacted each of them with the intention to finalise 8 schools for the case study analysis.

Effect size calculation

Effect sizes will be reported using Hedge's *g* (Cohen's *d* bias corrected) will be accompanied by 95% confidence intervals as per EEF guidelines.

References

Allen, R. (2016, September 1st). Consistency in Key Stage 2 writing across local authorities appears to be poor [Blog post]. Retrieved from http://educationdatalab.org.uk/2016/09/consistency-in-key-stage-2-writing-across-local-authorities-appears-to-be-poor/

Sun, J., & Stewart, D. (2007). Age and gender effects on resilience in children and adolescents. *International Journal of mental health promotion*, *9*(4), 16-25.

Report tables

MDES for Year 5

Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	140/5600 (70/2800; 70/2800)	0.7	0.14	80%	0.05	0.20
Randomisation	136/6338 (66/3027; 68/3311)	0.7	0.07	80%	0.05	0.20
Analysis (i.e. available pre- and post-test)				80%	0.05	

Balance at baseline – Year 5

Variable	Intervention group		Control group	
School-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
School Type:				
Academy Converted	14/66 (0)	21%	16/68 (0)	21%
Academy Sponsored	9/66 (0)	14%	7/68 (0)	13%
Community School	34/66 (0)	52%	30/68 (0)	50%
Foundation School	1/66 (0)	2%	2/68 (0)	1%
Voluntary Added School	5/66 (0)	8%	8/68 (0)	7%
Voluntary Controlled School	3/66 (0)	5%	5/68 (0))	4%
Ofsted rating:				
Outstanding	6/66 (0)	9%	7/68 (0)	10%
Good	48/66 (0)	73%	51/68 (0)	75%
Requires Improvement	10/66 (0)	15%	7/68 (0)	10%
Inadequate	2/66 (0)	3%	3/68 (0)	4%
Location/Setting:				
Urban	53/66 (0)	80%	56/68 (0)	78%
Rural	13/66 (0)	20%	1268 (0)	20%
School-level (continuous)	n (missing)	[Mean]	n (missing)	[Mean]
% of Pupil Eligible for FSM	66 (0)	18.2%	68 (0)	17.6%
% of SEN Pupil	66 (0)	14.3%	68 (0)	16.0%
% Reading, Writing and Mathematics Combined Attainment at KS2 4+	66 (0)	79.2%	68 (0)	79.6%
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
Eligible for FSM: Yes	882/3027 (0)	29%	973/3311 (0)	29%
Special Educational Needs: S, E or K	512/3027 (20)	17%	622/3311 (0)	19%
Gender: Male	1560/3027 (0)	52%	1700/3311 (0)	51%
Gender: Female	1467/3027 (0)	48%	1611/3311 (0)	49%
Attendance %				

Pupil-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Effect Size
KS1 reading point score	2836 (191)	15.49 (3.60)	3062 (249)	15.68 (3.66)	-0.05

KS1 mathematics point score	2836 (191)	15.92 (3.46)	3062 (249)	16.16 (3.47)	-0.07
KS2 reading score					
KS2 mathematics score					

Balance at baseline – student resilience survey (secondary outcome) – Year 5

Variable	Intervention group		Control group	
Pupil-level Survey Subscale: Self-Esteem (categorical)	**n/N (missing)**	**Average Response Rate**	**n/N (missing)**	**Average Response Rate**
Q1 - I can work out my problems	2813/3027 (214)		3084/3311 (227)	
Q2 - I can do most things if I try	2802/3027 (225)	92.8%	3073/3311 (238)	92.9%
Q3 - There are many things that I do well	2810/3027 (217)		3075/3311 (236)	
Variable	**Intervention group**		**Control group**	
Pupil-level Survey Subscale: Goals & Aspiration (categorical)	**n/N (missing)**	**Average Response Rate**	**n/N (missing)**	**Average Response Rate**
Q4 - I have goals and plan for the future	2796/3027 (231)	92.4%	3062/3311 (249)	92.4%
Q5 - I think I will be successful when I grow up	2795/3027 (231)		3054/3311 (257)	
Variable	**Intervention group**		**Control group**	
Pupil-level Survey Subscale: Home Connection (categorical)	**n/N (missing)**	**Average Response Rate**	**n/N (missing)**	**Average Response Rate**
Q6 - There is an adult at home who is interested in my	2789/3027 (238)		3059/3311 (252)	
Q7 - There is an adult at home who believes I will be a success	2797/3027 (230)	92.2%	3051/3311 (252)	92.3%
Q8 - There is an adult at home who wants me to do my best	2785/3027 (242)		3059/3311 (252)	
Q9 - There is an adult at home who listens to me when I have something to say	2789/3027 (238)		3057/3311 (254)	
Variable	**Intervention group**		**Control group**	
Pupil-level Survey Subscale: School Connection (categorical)	**n/N (missing)**	**Average Response Rate**	**n/N (missing)**	**Average Response Rate**
Q10 - There is an adult at home who is interested in my schoolwork	2787/3027 (240)		3058/3311 (253)	
Q11 - There is an adult at school who tells me when I have done well	2793/3027 (234)	92.0%	3057/3311 (254)	92.5%
Q12 - There is an adult at school who listens to me when I have something to say	2783/3027 (244)		3067/3311 (244)	
Q13 - There is an adult at school who believes I will be a success	2780/3027 (247)		3067/3311 (244)	

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
KS2 reading scores (complete cases)							
KS2 reading scores (multiple imputations)							

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
Attendance (complete cases)							
Attendance (multiple imputations)							
Resilience (complete cases)							
Resilience (multiple imputations)							
KS2 mathematics (complete cases)							
KS2 mathematics (multiple imputations)							

Appendix – Year 4 Data

Participant flow diagram for Year 4
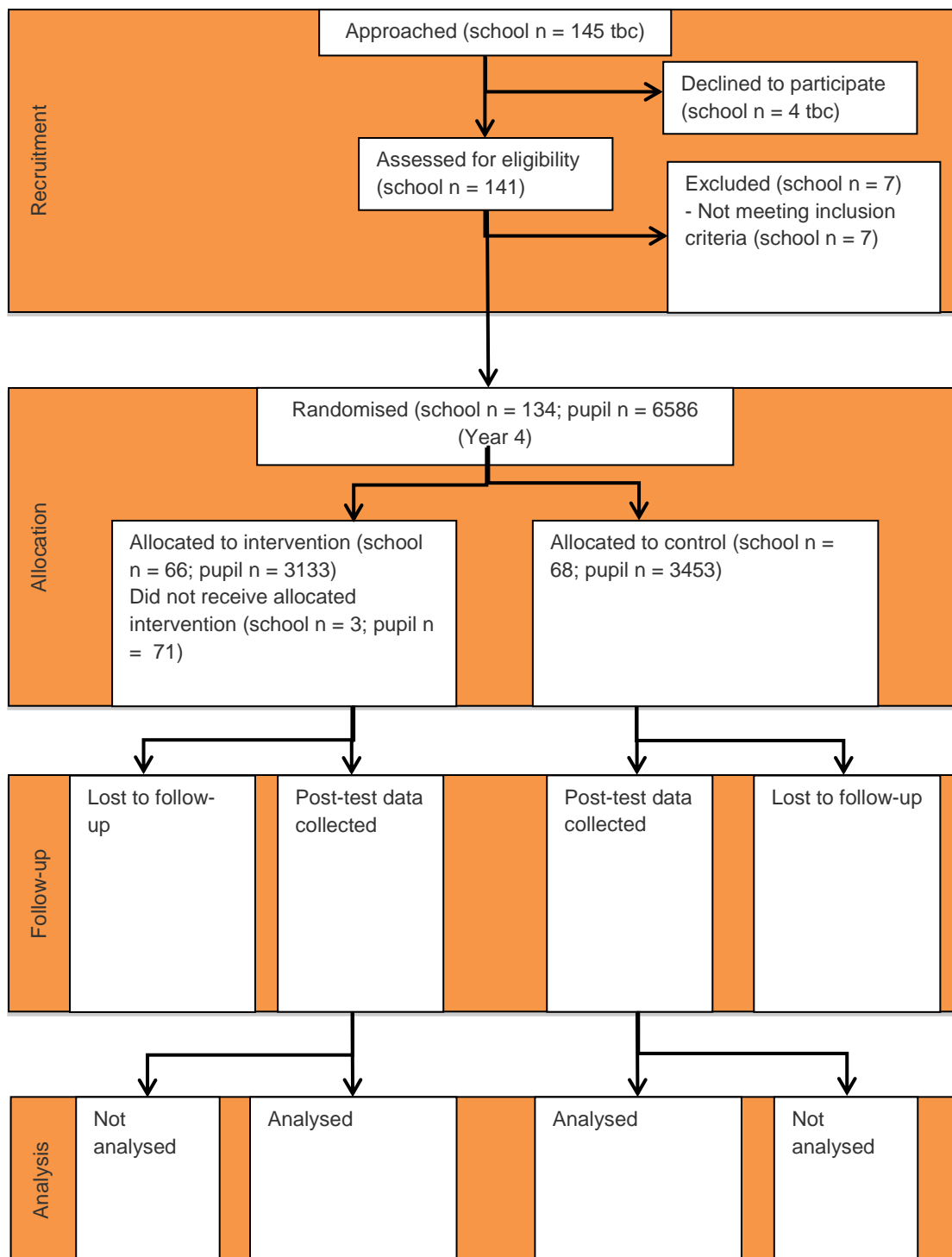
MDES calculation for Year 4 data set

In the actual Year 4 sample, a total of 6586 pupils over 134 schools were included. Thus, the power calculations for the randomised sample are repeated as follows:

Table 4 – MDES Estimates – Achieved Sample

Cohort	No. of clusters	Average cluster size (Year 4)	Pre-Post test correlation/R^2	ICC[8]	Power	Significance	MDES
Whole Cohort	134	49	0.7/0.49	0.07	0.8	0.05	0.12
FSM Cohort	134	14	0.7/0.49	0.07	0.8	0.05	0.18
AfA Target Cohort	134	10	0.7/0.49	0.07	0.8	0.05	0.20

---

[8] Calculated based on KS1 literacy point scores obtained from the NPD data.