

Achievement for All

Addendum Report

April 2021

Humphrey, N., Squires, G., Choudry, S., Byrne, E.,
Demkowicz, O., Troncoso, P. & Wo, L.

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.


The EEF aims to raise the attainment of children facing disadvantage by:


- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.


The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus (formerly Impetus Trust) and received a founding £125m grant from the Department for Education.


Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

 Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

 0207 802 1653

 jonathan.kay@eefoundation.org.uk

 www.educationendowmentfoundation.org.uk

Contents

About the evaluator	4
Executive summary	5
Introduction	7
Methods.....	8
Impact evaluation.....	13
Conclusion.....	23
References.....	25
Appendix 1: MLM ITT and subgroup analyses	27

About the evaluator

The project was independently evaluated by a team from the Manchester Institute of Education, University of Manchester led by Prof Neil Humphrey and Prof Garry Squires.

Dr Sophina Choudry and Dr Elizabeth Byrne were responsible for managing the trial, including data generation and analysis of both the RCT and IPE strands used in the main report. Dr Patricio Troncoso undertook additional quantitative analysis and completed the analysis for this addendum report. Dr Ola Demkowicz supported data generation in the IPE strand of the trial and led the analysis of the qualitative data used in the main report.

Lawrence Wo supported data management. All of the above-named staff worked for the University of Manchester during the trial.

Contact details:

Name: Professor Neil Humphrey

Address: Ellen Wilkinson Building, University of Manchester, Oxford Road, Manchester M13 9PL

Tel: +44 161 275 3404

Email: neil.humphrey@manchester.ac.uk

At the request of Achievement for All, the delivery team are not named individually in this report.

Executive summary

This addendum report should be read in conjunction to the main report of this project, Achievement for All: Evaluation Report (Humphrey et al., 2020). The main report presented the findings relating to the first cohort of 6,338 pupils who were in Year 5 at the start of the 2016/2017 academic year. A randomised controlled trial was used to compare outcomes in reading, maths, resilience-related outcomes, and attendance among pupils in 66 schools who received Achievement for All (AfA) with those who attended 68 schools who continued with business as usual. The main report also presents the implementation and process evaluation that was undertaken during the trial.

In brief, for the Year 5 cohort, we found that AfA led to a two-month reduction in reading progress at both the whole-group and subgroup levels—AfA target children and those eligible for free school meals ('FSM children'). ('AfA target children' are defined as those identified as 'vulnerable to underachievement'.) All children and FSM children in the Achievement for All schools made two months less progress in maths, on average, compared to equivalent children in schools that did not receive the programme, while AfA target children made three months less progress in maths, on average, compared to target children in control schools. However, children in AfA schools were more likely to report that there was an adult in their school who cared about them and supported them.

This addendum report focuses on the above-noted outcomes for 6,586 pupils who were in Year 4 at the start of the 2016/2017 academic year. This second cohort of pupils were exposed to AfA for the full six terms (two school years) recommended by the developer; this contrasts with the five terms of intervention exposure experienced by the first cohort in the main report. In both the main and addendum report, subgroup moderator analysis was undertaken for the vulnerable 'AfA target' group of pupils identified by the schools involved in the project in addition to the standard subgroup analysis pertaining to FSM children. The main report can be accessed [here](#).

The project

The intervention in brief

Achievement for All is a whole-school development programme that was developed from a national pilot funded by the Department for Education just over ten years ago. It aims to improve pupil academic outcomes such as reading and mathematics as well as resilience-related outcomes such as goals and aspirations and attendance. Schools are supported by a coach from the charity AfA 3As (Aspiration, Access, and Achievement) who works with them to develop leadership and governance, teaching and learning, parent and carer engagement, and wider outcomes and opportunities. The intervention is very flexible and driven by a needs analysis of the school that leads to an action plan to be implemented over a period of two years. Part of the needs analysis involves identifying a target group of pupils with an aim to address the academic achievement gap between the lowest 20% of children and their peers.

Impacts measured by the addendum

In this addendum report we consider the second cohort of pupils who were in Year 4 at the start of the 2016/2017 academic year and received the full two-year intervention (six terms). The findings are reported for reading—the primary outcome—and for secondary outcomes relating to maths, 'resilience' (self-esteem, goals and aspirations, family connection, and school connection), and attendance. Attainment and attendance data was obtained from the National Pupil Database in December 2016 and November 2019 and resilience-related outcome measures were collected using online surveys taken in September and October 2016 and April and May 2018.

Summary of previous results

The main report is concerned with Cohort 1/Year 5 for whom outcome measures (academic outcomes, resilience-related outcomes, and attendance) were taken after five terms of the AfA project running in the treatment arm schools. These pupils made significantly less progress than those in the control arm of the trial. Children in the AfA arm made two months less progress, on average, in reading. This was true for Cohort 1 as a whole, for the children identified as 'vulnerable to underachievement' ('AfA target' children), and for children eligible for FSM. Children eligible for FSM in the AfA schools made two months less progress in maths, on average, compared to FSM children in control schools (ES: -0.17). AfA target children in the intervention arm made three months less progress in maths on average compared to equivalent children in control schools (ES: -0.19). The AfA programme did not improve pupil's self-esteem, goals and

aspirations, perceptions of how supportive their families were, or the attendance of target children. However, children in AfA schools were significantly more likely to report that there was an adult in their school who cared about and supported them (ES: 0.15).

Summary of new results

Our analyses indicate that being exposed to the full six terms of the AfA intervention leads to broadly the same outcomes as those observed following five terms of exposure (see Table 1). There is a negative impact on the whole cohort and a negative impact on the target pupils for the primary outcome of reading attainment and for mathematics. AfA was not found to have a beneficial effect on resilience measures for the whole cohort; however, pupils eligible for FSM were more likely to report higher levels of goals and aspirations.

Table 1: Summary of impact on primary outcomes for the whole cohort and the AfA target group

Outcome/ Group	Effect size (95% confidence Interval)	Estimated months' progress	EEF security rating	No. of pupils	P value	EEF cost rating
Reading for whole group (Main Report, Y5 cohort)	-0.12 (-0.17, -0.07)	-2	5	5,813	0.008	£
Reading for whole group (Addendum Report, Y4 cohort)	-0.10 (-0.15; -0.05)	-2	N/A	6,074	0.016	N/A
Reading for AfA target group (Main Report, Y5 cohort)	-0.16 (-0.27, -0.05)	-2	5	1,231	0.026	£
Reading for AfA target group (Addendum Report, Y4 cohort)	-0.15 (-0.26; -0.04)	-2	N/A	1,224	0.030	N/A

Key conclusions

1. Children in the Achievement for All schools made two months less progress in reading, on average, compared to children in schools that did not receive the programme.
2. Target children in the Achievement for All schools (the lowest 20% of attainers or those deemed to be 'vulnerable to underachievement' as identified by their school) made two months less progress in reading, on average, compared to target children in schools that did not receive the programme. Children eligible for FSM in Achievement for All schools made one month less progress in reading, on average, compared to FSM-eligible children in schools that did not receive the programme.
3. All children, AfA target children, and FSM-eligible children in the Achievement for All schools made two months less progress in maths, on average, compared to equivalent children in schools that did not receive the programme.
4. The evaluation found that the programme did not improve pupils' self-esteem, goals and aspirations, perceptions of how supportive their families were, perceptions of how supportive their schools were, or target children's attendance. However, FSM-eligible children in Achievement for All schools were more likely to report higher levels of goals and aspirations.
5. The findings noted above are largely consistent with those documented in the main AfA trial report.

Introduction

Intervention

The Achievement for All programme (AfA) was developed from the National Pilot for Achievement for All (Humphrey and Squires, 2010, 2011a, 2011b) and is a whole-school improvement programme that aims to improve pupil outcomes in academic performance (such as reading and mathematics), resilience (self-esteem, goals and aspirations, family connection, and school connection), and attendance.

AfA was originally designed to focus on the lowest-achieving 20% of pupils in a school; in the national pilot, these were children with special educational needs and disabilities (SEND). In the pilot programme, there were three main strands and the schools involved had a high level of support from local authorities, national strategies advisors, and the National College for School Leadership—and they had additional funding. Schools were involved in the national pilot for two years and worked on assessment, tracking, and intervention, structured conversations with parents, and provision for developing wider outcomes (such as improved attendance). The pilot was conceptualised as requiring strong leadership and clear school action plans that were then implemented throughout the duration of the project. Statistically significant improvements were found for children with SEND in English and mathematics, with effect sizes considered large enough to be practically meaningful (Barlow et al., 2015). Attendance improved for those pupils who were persistent absentees at the start of the pilot, parental engagement improved over the pilot, and teachers reported improvements in peer relationships and reductions in bullying and behaviour problems compared to a control group of pupils with SEND from schools not implementing AfA (Humphrey et al., 2013).

For the current programme, a national charity was formed in 2011 called AfA 3As (Aspiration, Access and Achievement). Schools subscribe to the AfA programme using their existing funding and a coach from the charity works with them. The three main areas of the original pilot programme remain but have been renamed and further developed, and leadership and governance is now a distinct core element. A randomised controlled trial (RCT) of the revised version of AfA was conducted and found statistically significant negative impacts on attainment at both the intent to treat (ITT) and subgroup (FSM pupils) levels of analysis (Churches, 2016). However, both the national pilot and the Churches study had numerous limitations. The pilot evaluation was not an RCT, relied on teacher assessment of academic outcomes, and examined AfA under ideal conditions that could not be replicated at scale. While the Churches study used an RCT design and tested the newer version of the programme, it suffered from substantial attrition that rendered the security of findings very weak.

Our EEF-funded evaluation involved an RCT in which primary schools were randomly allocated to the intervention arm (AfA) or control arm (business as usual). The RCT was registered with the ISCRTN (registry trial reference number ISRCTN67347514, details here). There were two cohorts of pupils in Year 5 and Year 4 (aged eight to ten years). Those pupils in Year 5 (Cohort 1) were either attending schools that received five terms of AfA or they were in control schools. Pupils in Year 4 (Cohort 2) were in the same schools as Cohort 1 but received six terms of AfA (or business as usual) in order to mimic the length of time that AfA would normally be in place (and replicate the dosage conditions of the pilot).

The main report (Humphrey et al., 2020) describes the intervention in more detail and reports on the implementation and process evaluation. It also describes the RCT for Cohort 1, for which outcome measures (academic outcomes, resilience-related outcomes, and attendance) were taken after five terms of the AfA project running in the treatment arm schools. These pupils made significantly less progress than those in the control arm of the trial:

- Children in the AfA arm made two months less progress, on average, in reading compared to children in the control arm who did not receive the programme (effect size, ES: -0.12).
- Children identified by the AfA schools as 'vulnerable to underachievement' (referred to as 'AfA target' children) made two months less progress in reading, on average, compared to their counterparts in the control schools (ES: -0.16). Children eligible for FSM in AfA schools made two months less progress, on average, in reading compared to FSM children in control schools (ES: -0.12).
- All children (ES: -0.11) and those eligible for FSM (ES: -0.17) in the AfA schools made two months less progress in maths, on average, compared to their counterparts in control schools. AfA target children in the intervention arm made three months less progress in maths, on average, compared to equivalent children in control schools (ES: -0.19).
- The AfA programme did not improve pupil's self-esteem, goals and aspirations, perceptions of how supportive their families were, or the attendance of target children. However, children in AfA schools were significantly

more likely to report that there was an adult in their school who cared about them and supported them (ES: 0.15).

Evaluation objectives

It was evident in the national pilot that some schools experienced a slow start to understanding the flexibility of AfA and progress only became evident after the first year (Humphrey and Squires, 2010). One potential explanation for the negative impact on Cohort 1 pupils noted in our main report was that they had not received the full intervention (Humphrey et al., 2020). To explore this possibility, this addendum report focuses on outcomes for Cohort 2, who received a full six terms of the AfA intervention (Jan–Mar 2017, Apr–Jul 2017, Sep–Dec 17, Jan–Mar 18, Apr–Jul 18, Sep–Dec 18; outcomes assessed May 19). Our research questions are:

RQ1. Compared to usual practice, what is the impact of AfA on children’s literacy (primary outcome), maths, attendance, and resilience-related outcomes¹ (secondary outcomes) after six terms of exposure?

RQ2. In relation to RQ1 above, are there differential intervention benefits in the above outcomes among pre-specified subgroups of children?

- A. Among children eligible for free school meals?
- B. Among the target group of children identified by participating schools as belonging to ‘the lowest achieving 20%’ (referred to as ‘AfA target’ children)?

The evaluation protocol, statistical analysis plan, and main report can all be found [here](#). The latter provides details regarding ethical approval and data protection processes, including the legal basis for processing data, for the project as a whole (including the data discussed in this addendum report).

Methods

Trial design

Trial type and number of arms		Two-arm RCT, school as unit of randomisation
Unit of randomisation		Schools
Stratification variable(s) (if applicable)		%FSM, %SEN, %RWM
Primary outcome	Variable	Reading attainment (Y4 cohort and Y4 AfA target group) (regarded as co-primary outcomes)
	Measure (instrument, scale)	Key Stage 2 (KS2) English reading marks (marks for reading only)
Secondary outcome(s)	Variable(s)	Mathematics attainment (Y4 cohort) Attendance of AfA target pupils (Y4 cohort) Resilience related outcomes (Y4 cohort)
	Measure(s) (instrument, scale)	KS2 mathematics marks Number of unauthorised absences Subscales of the Student Resilience Survey (SRS): self-esteem, goals and aspirations, school family connection, and school connection
Baseline for primary outcome	Variable	Reading and writing attainment (Y4 cohort and Y4 AfA target group)
	Measure (instrument, scale)	KS1 English reading and writing marks (combined)
Baselines for secondary outcomes	Variable(s)	Reading and writing attainment (Y4 cohort) Mathematics attainment (Y4 cohort) Attendance of AfA target pupils (Y4 cohort) Resilience related outcomes (Y4 cohort)
	Measure(s) (instrument, scale)	KS1 mathematics marks Number of unauthorised absences Subscales of the Student Resilience Survey (SRS): self-esteem, goals and aspirations, school family connection, and school connection

¹ Children’s self-reported self-esteem, goals and aspirations, family connection, and school connection.

A two-arm cluster RCT design was used. School was used as the unit of randomisation. Other forms of randomisation (at, for example, class or year level) were not feasible as AfA is a whole-school intervention. Schools were randomly allocated to deliver the AfA programme (intervention arm) or to a business as usual (control) arm. See the Randomisation section of the main report for more details. In order to minimise differential attrition by trial arm, control schools received a retention incentive of £1,000 (paid in instalments: £200 following random allocation, £200 at the end of the first year of the trial, £200 at the midpoint the second year of the trial, and £400 at the conclusion of the trial and on completion of required data/surveys). Schools allocated to the intervention arm were trained and instructed to implement the AfA programme during the two-year period (academic years 2016/2017 and 2017/2018).

Participant selection and sample size

Prospective trial schools were required to not already be, or have previously been, implementing the AfA programme. Schools were recruited by AfA 3As and were located in 78 of the 343 local authorities in England. The recruitment strategy to identify prospective schools included 'talking head' videos, presentations, and network events.²

In total, 145 schools were approached. Of these schools, four declined to take part and seven were excluded for not meeting the eligibility requirements. This resulted in 134 schools being recruited to the trial, leading to them signing the MoA; 66 schools were allocated to the AfA intervention arm and 68 to the control arm (see the Randomisation section for more details). The target cohort in this addendum report were pupils in Year 4 (Y4) in the first year of the trial (2016/2017). After accounting for parental/carers opt-outs (n = 63; 0.95%), this sample consisted of n = 6,586 pupils. Within the cohort, 1,350 (21%) were nominated by participating schools as the AfA target cohort. The nomination process was undertaken jointly by schools and coaches as part of the normal AfA procedure based on joint decision-making.

Addendum outcome measures

Primary outcome measure

Academic attainment in reading

The primary outcome measure for the trial was pupils' academic attainment in reading. Data was sourced from the NPD at baseline (pre-trial, summer 2015) and at the conclusion of the trial (post-trial, summer 2019) for pupils in the Y4 cohort. End of KS1 literacy scores (the 'KS1_READWRITPOINTS' variable) were used as the pre-trial covariate and end of KS2 reading scores (the 'KS2_READMRK' variable) were used as the main post-trial outcome.

Secondary outcome measures

Academic attainment in mathematics

Academic attainment scores in mathematics were also assessed. Pre-trial KS1 scores (the 'KS1_MATPOINTS' variable, summer 2015) were entered into models as the baseline covariate with KS2 scores as the outcome measure (the 'KS2_MATMRK' variable, summer 2019).

Resilience-related outcomes

Subscales of the Student Resilience Survey (SRS; Sun and Stewart, 2007) were delivered via a secure online survey platform (World App Key Survey). The items in this survey are available in Appendix F of the main report. The survey was used to assess pupils' self-reported protective factors, including their ratings of 'self-esteem' (three items), 'goals and aspirations' (two items), 'family connection' (four items), and 'school connection' (four items). These four areas were chosen during discussions between the University of Manchester, AfA, and the EEF as being those that provided the best measure of the non-academic outcomes noted in the AfA theory of change. Pupils responded to a series of statements, such as 'I can do most things if I try', on a five-point scale (with a score of one corresponding to 'never', and a response of five meaning 'always'). The subscales have good internal consistency (that is, items correlated well with

² Initially schools were contacted by AfA 3As via an email campaign, which was followed up with phone calls to clarify details regarding the programme and the trial. Regional leads at AfA 3As followed-up with further details and made contact to discuss the start-up process.

one another) and are also negatively associated with mental health problems, demonstrating the validity of the SRS as a measure of resilience-related factors (Lereya et al., 2016).

Attendance

Attendance data was also assessed, however, given the uniformly high attendance rates across primary schools, analyses of this outcome are restricted to pupils in the AfA target group. Absence data for the whole academic year at pre-trial (2015/2016) and during the final year of the trial (2018/2019) was obtained for the members of this subgroup in the Y4 cohort.

For this measure, the initial plan was to calculate the percentage of half-days missed due to unauthorised absence using two variables provided by the NPD, namely the number of overall absences (for the whole academic year, the 'SessionsPossible_6HalfTerms_ab16' and 'SessionsPossible_6HalfTerms_ab18' variables) and the number of sessions possible (the 'UnauthorisedAbsence_6HalfTerms_ab16' and 'UnauthorisedAbsence_6HalfTerms_ab18' variables). This would then be coded into a binary variable of scores being < 10% or ≥ 10%.³ When the SAP was written there was an assumption, due to the nature of the AfA target group, that there would be a reasonable proportion of pupils identified as persistent absentees. However, as noted in the main report, after accounting for missing data, only 2.6% and 3.2% were categorised as persistent absentees for the pre-trial and post-trial scores, respectively. Therefore, also for consistency with the main report, we used count data as our outcome measure for the attendance variable (that is, the number of unauthorised absences for the 2018/2019 academic year) and our analysis was amended accordingly (see the Deviations from the SAP section below).

Academic attainment in writing, and reading and writing combined

Due to recent changes in the way that writing is assessed at Key Stage 2, we have not conducted the planned analyses involving this measure. This is due to writing being teacher-assessed, with a potential to introduce bias. This decision is consistent with current EEF policy in relation to the use of KS2 writing data.

Statistical analysis

To address RQ1, ITT complete case analyses were undertaken for all primary and secondary outcome measures using raw complete case data. In ITT, participants' data is analysed according to the group to which they were randomly assigned, irrespective of what happened after the randomisation process was completed (for example, partial implementation). We used multilevel models (MLM) with fixed slopes and random intercepts for all outcome measures apart from attendance. On account of the nested nature of the data, models with two levels (pupils, clustered in schools) were fitted, controlling for baseline (pre-test) scores at the pupil level. Trial group (AfA intervention versus business as usual) was entered as a school-level predictor and post-test scores were used as the response variable in each case (Model 1.1).

For RQ2A (FSM eligible children) and RQ2B (AfA target group), subgroup analyses were performed for each of the primary and secondary outcome measures. Each model described above was re-run twice with a subset of the main data. In the first subgroup analysis, only FSM pupils were included in the analysis (Model 2.1), and in the second subgroup analysis, only AfA target pupils were entered (Model 2.2).

In line with EEF guidelines, further sensitivity tests were conducted. First, a number of explanatory variables were added to Model 1.1 for each outcome measure (resulting in Model 1.2). At the school level, the minimisation variables (%FSM, %SEN, and %RWM+4, as per EEF guidelines) and the usual practice indicator⁴ were entered as co-variates. At the pupil level, SEN and gender were added as co-variates.

All the models described above were estimated using Stata (version 16.1; see the Deviations from the SAP section below for more information).

Next, for outcome measures where the proportion of incomplete cases exceeded 5% (see the Missing Data section), multiple imputation (MI) procedures were used to re-estimate models to deal with missing data (whereby partially

³ Whereby a threshold of 10% was to be applied as per the Department for Education's current definition of persistent absence. Accordingly, pupils with an absence rate of 10% or more are classed as persistent absentees.

⁴ Derived from the Usual Practice Survey (UPS) to provide a more robust estimate of the achieved relative strength of AfA.

observed cases could be included in the analysis). In these cases, the previously described models, including the primary ITT analysis (Model 1.1), subgroup analyses (Models 2.1 and 2.2), and the analysis with the additional co-variables (Model 1.2) were repeated using MI. For all the outcome measures, models were estimated using joint modelling MI, which was implemented in R (version 3.6.1) using the package jomo (version 2.6-9; see the Deviations from the SAP section below for more information).

For all models in the primary, secondary, and sensitivity analyses, a statistically significant trial group coefficient (that is, where $p < 0.05$) was used to determine whether the null hypothesis could be rejected (for example, intervention effects were present). Hedge's g (Cohen's d bias corrected; Hedges, 2007) effect sizes were calculated (see the Effect Size Calculations' section below) along with 95% confidence intervals, as per EEF reporting guidelines. The log-likelihood of models and changes in variance partition coefficient (VPC) between models are reported. Exact p values are reported for these analyses, enabling the reader to consider for themselves whether a given effect estimate would be statistically significant at a different Alpha value (for example, if opting to correct for multiple comparisons using the Bonferroni method).

Deviations from the SAP

For the attendance measure, the statistical analysis deviated from the SAP for the reasons outlined in the Attendance subsection above. We conducted a multilevel negative binomial regression using count data in Stata (version 16.1). The response variable of post-trial absence was modelled with trial group being entered as a level two predictor and pre-trial absence being added as a level one explanatory variable (Model 1.1). For the sensitivity analyses, we then added the additional explanatory variables simultaneously to the model (Model 1.2).

Planned statistical tests that modelled the outcome measures of KS2 writing and KS2 reading and writing combined were not conducted for the reasons outlined in the Academic Attainment in Writing, and Reading and Writing Combined subsection above.

A further deviation from the SAP involved the statistical software programmes used to conduct the MLM analyses. All models using listwise deletion were estimated using Stata (version 16.1). This deviation is due to the ONS not being able to provide access to MPlus software on the Secure Research Service (SRS). For the resilience outcome measures, the plan was to run the models in Mplus; however, the FSM subgroup analysis required NPD data, which is why they had to be conducted in the SRS, where Mplus is unavailable. Additionally, all models were re-estimated using joint modelling multilevel multiple imputation, as implemented by the R (version 3.6.1) package jomo (version 2.6-9) to account for missing data due to the FIML method being unavailable in the Secure Research Service.

Effect size calculations

Effect sizes were calculated using Hedge's g formula (Hedges, 2007):

$$g = J * \frac{(\bar{x}_1 - \bar{x}_2)_{adjusted}}{s^*}$$

Where $(\bar{x}_1 - \bar{x}_2)_{adjusted}$ denotes the different in means between the trial groups adjusting for pre-test score, and was retrieved from the coefficient of the trial group effect of the ITT model. The pooled SD (s^*) and Hedge's bias correction (J) were calculated as follows:

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}} \quad J = \left(1 - \left(\frac{3}{4(n_1 + n_2) - 9}\right)\right)$$

Where n_1 and s_1 correspond to the n and SD of the control group, respectively, and n_2 and s_2 represent the n and SD of the intervention group, respectively.

Missing data

We calculated the proportion of missing data for the intervention and control arms for the primary outcome measure. Missing observations were due to failure of the NPD to match data and other reasons such as pupil absence on the day of the tests/surveys, or incomplete tests/surveys. For outcome measures where the extent of missing cases exceeded

5%, we performed additional sensitivity analyses by re-estimating the statistical models using multilevel multiple imputation in the R (version 3.6.1) package jomo (version 2.6-9).

We also examined missing data by conducting a regression analysis with a binary variable for complete (0) and incomplete (1) cases. A complete case was defined by post-test (KS2 reading) scores being available. The binary variable was entered into the regression analysis as the outcome variable, with the baseline measure (KS1 reading and writing), condition (treatment or usual practice), FSM (if eligible), and AfA target cohort (if AfA target) entered as explanatory variables, alongside relevant interactions between them. This allowed us to determine the likelihood of pupils having complete cases, and whether this is influenced by trial group allocation (treatment versus control), or subgroup membership (for example, FSM eligibility, AfA target group).

Impact evaluation

Participants

One hundred and thirty-four schools (6,586 pupils in Cohort 2/Y4) were recruited for the trial. Of this sample, 66 schools (3,133 pupils) were allocated to the intervention arm, and 68 (3,453 pupils) to the control (business as usual) arm. For the primary analysis, pre-trial baseline data (KS1 reading and writing combined points) was available for 6,214 pupils (94%). At post-trial, outcome data (KS2 reading mark) was available for 6,218 pupils (94%). Missing cases were due to absence of the day of test or incomplete tests, or the lack of a match in the NPD. Complete data was available for 6,074 pupils (92%). The sample size of our complete case analysis (Model 1.1) yielded power for an MDES of 0.11 (see Table 2 below). To note is that some values differ from the SAP because those presented in Table 2 are derived from Cohort 2/Y4, whereas those in the SAP are based on Cohort 1/Y5. Figure 1 depicts the flow of Y4 participants through the trial.

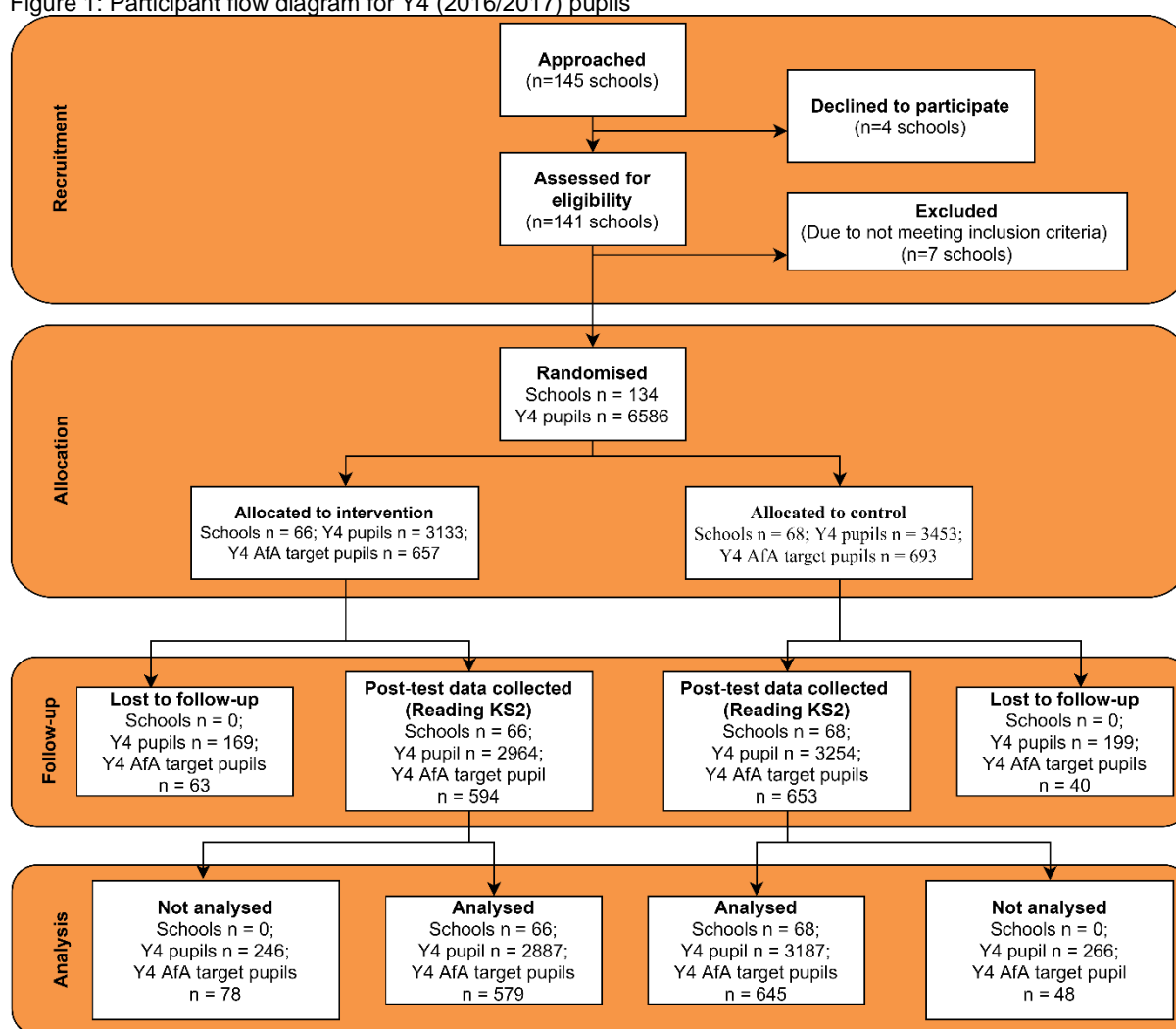
Table 2: Minimum detectable effect size at different stages of the trial—primary outcome

		Protocol			Randomisation			Analysis		
		Overall	FSM	AfA target	Overall	FSM	AfA target	Overall	FSM	AfA target
MDES		0.15	0.18	0.20	0.137	0.166	0.172	0.112	0.168	0.191
Pre-post test correlation (R)		0.49	0.49	0.49	--	--	--	--	--	--
R-squared		0.7	0.7	0.7	--	--	--	--	--	--
R^2_1	Level 1 (pupil)	--	--	--	0.474	0.414	0.414	0.442	0.394	0.293
R^2_2	Level 2 (school)	--	--	--	0.156	0.031	0.031	0.337	-0.278	0.282
ICC	Level 2 (school)	0.14	0.14	0.14	0.083	0.075	0.075	0.063	0.045	0.122
Alpha		0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Power		0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
One- or two-sided?		Two	Two	Two	Two	Two	Two	Two	Two	Two
Average cluster size		40	12	8	49.1	12	10.1	45.3	14	9.1
Proportion randomised to treatment		0.5	0.5	0.5	0.476	0.5	0.487	0.475	0.486	0.473
Number of schools	Intervention	70	70	70	66	66	66	66	66	66
	Control	70	70	70	68	68	68	68	68	68
	Total	140	140	140	134	134	134	134	134	134
Number of pupils	Intervention	2800	840	560	3133	804	657	2887	913	579
	Control	2800	840	560	3453	804	693	3187	964	645
	Total	5600	1680	1120	6586	1608	1350	6074	1877	1224

Note: MDES = minimum detectable effect size; ICC = intraclass correlation coefficient; R^2_1 = proportion of variance explained by baseline covariate at level 1; R^2_2 = proportion of variance explained by baseline covariate at level 2.

The MDES calculations at the protocol stage were carried out using the Optimal Design programme (version 3.01). The MDES calculations at the randomisation and analysis stages used the ‘mdesapp’ online calculator (Troncoso, 2020),⁵ which implements the MDES formula as described in Bloom et al. (2007) and the pre-testing guidelines of the Education Endowment Foundation (EEF, 2013). This table contains all the information necessary to replicate the MDES estimates in both software packages. ICC values vary at the protocol and randomisation stages because of the source of the data. The protocol and randomisation stages have used a combination of national KS1 data and the FRIENDS EEF trial evaluation (Wigelsworth et al., 2018). KS1 national data was used as the benchmark ICC for all cohorts at the protocol stage and the AfA target cohort of the randomisation stage. The FRIENDS data was used as a benchmark for the variance explained by the baseline covariate at levels 1 and 2 (R_1^2 and R_2^2) for all cohorts at the randomisation stage. The FSM and AfA cohorts have the same underlying model as benchmark for the explanatory power of the baseline. The FSM subgroup MDES estimate at randomisation stage uses the same hypothetical sample and cluster sizes as the protocol stage because the true FSM eligibility status is only known at the analysis stage in the ONS Secure Research Service (hence there are further marginal differences from the SAP, which was finalised at an earlier stage of the project).

Figure 1: Participant flow diagram for Y4 (2016/2017) pupils



In Figure 1 above, post-test data is KS2 scores, which do not align fully with KS1 (pre-test data) in terms of missing values. Analysed data (n in our main ITT model) comprises all those with complete KS2 and KS1 scores, so ‘not analysed’ are those with missing values in either KS2 or KS1. Thus, ‘analysed’ plus ‘not analysed’ reflects total n at allocation.

⁵ The online calculator is available at: <https://patricio-troncoso.shinyapps.io/mdesapp/>. Accompanying documentation can be found here: https://rpubs.com/patroncos/mdesapp_calculation. Source code can be found here: <https://github.com/patroncos/mdesapp>

Attrition

Attrition was 0% at the school level and 7.77% at the pupil level for Cohort 2/Y4 (7.85% and 7.7% for the intervention and control arms of the trial, respectively) based on a ratio of 6,586 randomised to 6,074 analysed for the primary trial outcome. This attrition rate is marginally lower than that documented in the main report for Cohort 1 (Y5: 8.28%).

Table 3: Pupil-level attrition from the trial—primary outcome, Y4 reading, whole group

		Intervention	Control	Total
N pupils	Randomised	3133	3453	6586
	Analysed	2887	3187	6074
Pupil attrition (from randomisation to analysis)	Number	246	266	512
	Percentage	7.85%	7.7%	7.77%

Table 4: Pupil-level attrition from the trial—primary outcome, Y4 reading, AfA target group

		Intervention	Control	Total
N pupils	Randomised	657	693	1350
	Analysed	579	645	1224
Pupil attrition (from randomisation to analysis)	Number	78	48	123
	Percentage	11.87%	6.92%%	9.33%

Tables 5 and 6 depict pupil-level characteristics for Cohort 2/Y4 (note that school-level characteristics are not reported as these can be found in the main report).

Table 5: Balance at baseline for the whole Y4 cohort

Outcome	Raw means				Effect size
	Intervention group		Control group		
	N = 3133		N = 3453		
n (missing)	Mean (SD)	n (missing)	Mean (SD)		
KS1 English	2952 (181)	16.629 (3.814)	3262 (191)	16.537 (3.782)	0.024
KS1 maths	2952 (181)	16.381 (3.547)	3263 (190)	16.232 (3.389)	0.043
Self-esteem	2876 (257)	11.543 (1.934)	3172 (281)	11.573 (1.924)	-0.016
Goals and aspirations	2876 (257)	7.981 (1.779)	3162 (291)	8.031 (1.762)	-0.028
Family connection	2756 (377)	17.414 (2.676)	3066 (387)	17.503 (2.637)	-0.033
School connection	2789 (344)	17.164 (2.844)	3040 (413)	17.341 (2.655)	-0.064

Table 6: Balance at baseline for the AfA target group Y4 cohort

Outcome	Raw means				Effect size
	Intervention group		Control group		
	N=657		N=693		
n (missing)	Mean (SD)	n (missing)	Mean (SD)		
KS1 English	618 (39)	13.796 (3.926)	673 (20)	13.865 (3.683)	-0.018
KS1 maths	618 (39)	13.893 (3.488)	673 (20)	13.984 (3.205)	-0.027
Self-esteem	594 (63)	11.037 (2.056)	621 (72)	10.913 (2.058)	0.060
Goals and aspirations	594 (63)	7.699 (1.876)	617 (76)	7.778 (1.958)	-0.041
Family connection	570 (87)	16.709 (3.080)	604 (89)	16.747 (3.021)	-0.012
School connection	584 (73)	16.998 (2.972)	592 (101)	16.993 (2.962)	0.002

These table indicate an excellent balance at baseline in key observables, both for the whole sample and the AfA target group (all ES < 0.1). This sets a robust foundation for the internal validity of the trial because it indicates successful randomisation and means that post-test differences can be more confidently attributed to the intervention.

Table 7: Multilevel binary logistic regression model to investigate missingness in the primary outcome.

		Missingness model		
		Intercept (SE) = 1.002 (0.500)		
Level		Coef.	SE	<i>p</i>
Pupil				
	KS1 score (baseline)	-0.417	0.042	0.000
	FSM eligible	0.989	0.570	0.083
	AfA target	0.147	0.596	0.805
School				
	Trial group (if AfA)	-0.151	0.573	0.792
	Variance	0.902	0.349	--
Interactions				
	KS1 score*Trial group (if AfA)	0.009	0.048	0.846
	KS1 score*FSM eligible	-0.001	0.049	0.984
	KS1 score*AfA target	-0.024	0.054	0.650
	Loglikelihood	-433.86308		
	VPC	0.215		
	n	6207		

The model indicates that missingness is driven mainly by KS1 scores at baseline, regardless of trial and target group allocation. This reduces a potential source of bias in our analysis, that of differential attrition. Likewise, FSM eligibility does not seem to have a meaningful impact on missingness in the outcome variable. There is also no significant impact of trial arm on attrition.

Analysis

Results

Table 8: Means and SDs of pupil outcomes at pre-test (baseline) and post-test whole Y4 cohort.

Outcome	Intervention group		Control group	
	Pre-test mean (SD)	Post-test mean (SD)	Pre-test mean (SD)	Post-test mean (SD)
KS1 English	16.629 (3.814)	--	16.537 (3.782)	--
KS2 reading	--	32.887 (9.846)	--	33.740 (9.830)
KS1 maths	16.381 (3.547)	--	16.232 (3.389)	--
KS2 maths	--	76.133 (23.877)	--	78.065 (23.699)
Self-esteem	11.543 (1.934)	11.731 (1.809)	11.573 (1.924)	11.746 (1.839)
Goals and aspirations	7.981 (1.779)	8.372 (1.678)	8.031 (1.762)	8.302 (1.745)
Family connection	17.414 (2.676)	18.063 (2.234)	17.503 (2.637)	17.947 (2.295)
School connection	17.164 (2.844)	17.271 (2.885)	17.341 (2.655)	17.144 (2.987)

Table 9: Means and SDs of pupil outcomes at pre-test (baseline) and post-test—Y4 AfA target cohort subgroup

Outcome	Intervention group		Control group	
	Pre-test mean (SD)	Post-test mean (SD)	Pre-test mean (SD)	Post-test mean (SD)
KS1 English	13.796 (3.926)	--	13.865 (3.683)	--
KS2 reading	--	25.981 (10.581)	--	27.435 (10.410)
KS1 maths	13.893 (3.488)	--	13.984 (3.205)	--
KS2 maths	--	58.462 (26.302)	--	62.130 (26.570)
Self-esteem	11.037 (2.056)	11.248 (1.846)	10.913 (2.058)	11.139 (1.839)
Goals and aspirations	7.699 (1.876)	8.182 (1.689)	7.778 (1.958)	8.031 (1.841)
Family connection	16.709 (3.080)	17.756 (2.459)	16.747 (3.021)	17.348 (2.789)
School connection	16.998 (2.972)	17.335 (2.878)	16.993 (2.962)	17.126 (3.096)
Absence ⁶	1.62% (3.036)	1.90% (4.372)	1.27% (2.509)	1.55% (2.867)

⁶ Percentage of possible sessions that a pupil was absent.

Figure 2: Histograms of KS2 (post-trial) reading marks by trial group. Panels A and B show the distribution of scores for the intervention and control groups, respectively.

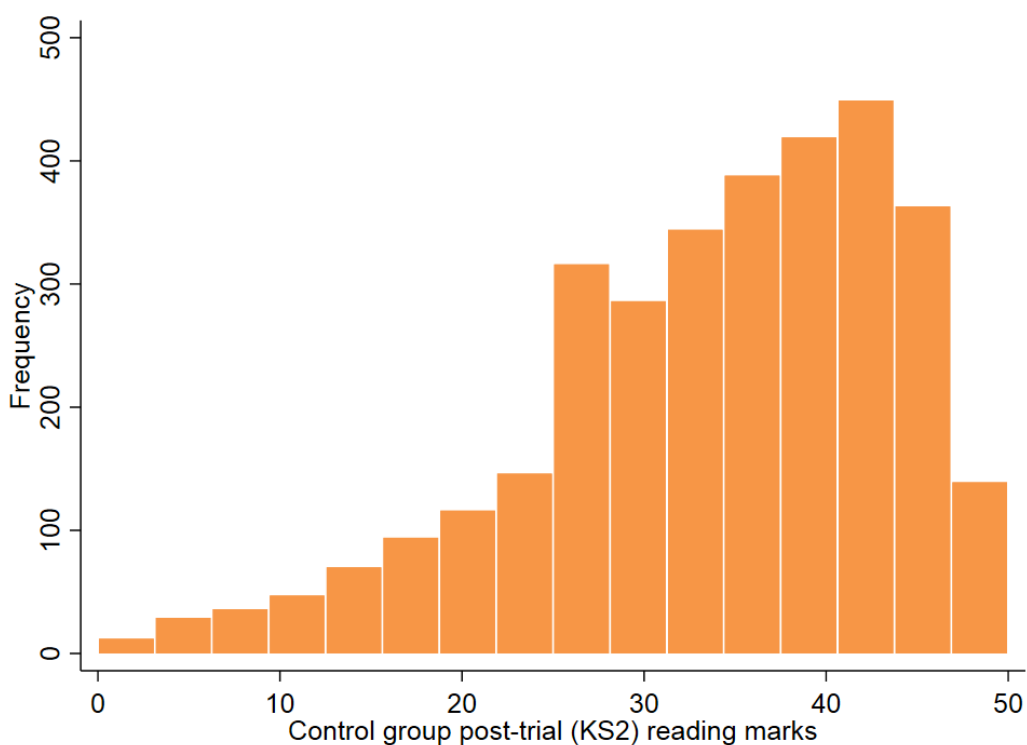
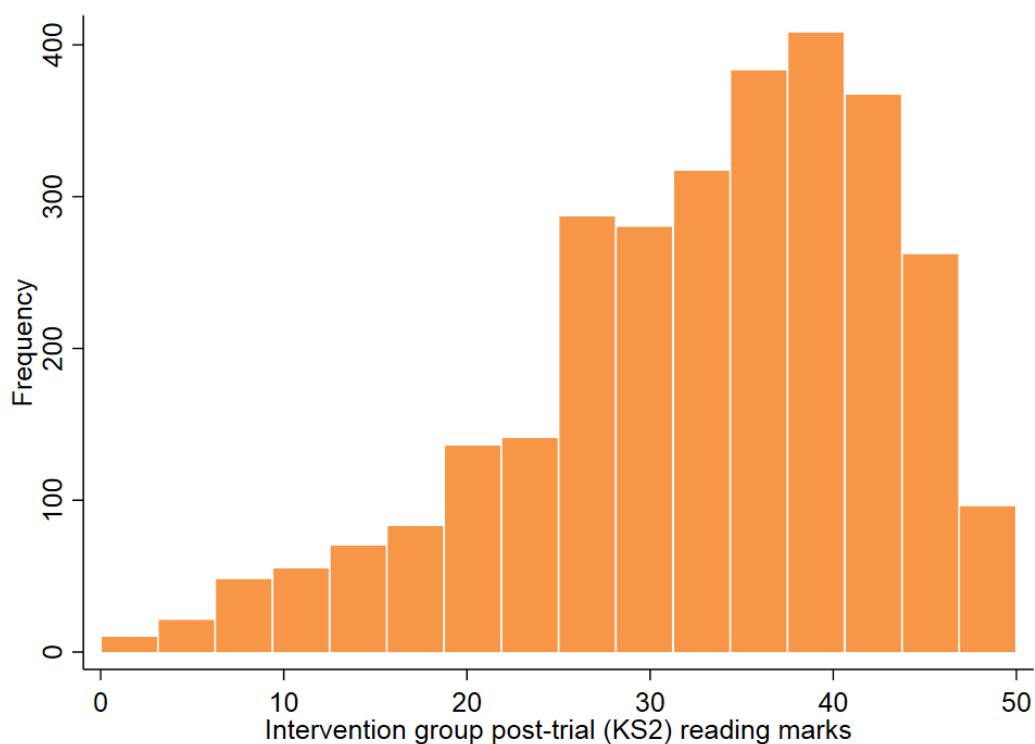


Table 10 provides a summary of the ITT and subgroup analyses. Full complete case models for all multilevel analyses are presented in Appendix 1 along with the results of additional sensitivity analyses (for example, added covariates and MI analyses accounting for missing data).⁷

⁷ Where applicable, that is, only for measures where the extent of missing cases exceeded 5%.

Table 10: Primary analyses—complete case

Outcome	Raw means				Effect size		
	Intervention group			Control group	n in model (intervention; control)	Hedges g (95% CI)	p-value
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)			
Main effects of intervention—ITT (Model 1.1)							
Reading	2964 (169)	32.887 (32.880; 32.893)	3254 (199)	33.740 (33.734; 33.746)	6074 (2887; 3187)	-0.096 (-0.146; -0.046)	0.016
Maths	2971 (162)	76.133 (76.117; 76.149)	3254 (199)	78.065 (78.051; 78.079)	6082 (2894; 3188)	-0.128 (-0.178; -0.078)	0.009
Self-esteem	2322 (811)	11.731 (11.730; 11.732)	2404 (1049)	11.746 (11.745; 11.747)	4443 (2178; 2265)	0.002 (-0.055; 0.059)	0.969
Goals and aspirations	2334 (799)	8.372 (8.371; 8.373)	2397 (1056)	8.302 (8.301; 8.303)	4442 (2193; 2249)	0.063 (0.006; 0.120)	0.144
Family connection	2277 (856)	18.063 (18.062; 18.065)	2329 (1124)	17.947 (17.945; 17.948)	4171 (2043; 2128)	0.061 (0.003; 0.118)	0.180
School connection	2288 (845)	17.271 (17.269; 17.273)	2345 (1108)	17.144 (17.142; 17.146)	4216 (2087; 2129)	0.088 (0.031; 0.146)	0.068
Subgroup effects of intervention—pupils eligible for FSM (Model 2.1)							
Reading	946 (17)	30.057 (30.036; 30.078)	990 (25)	30.318 (30.298; 30.338)	1866 (907; 959)	-0.089 (-0.178; 0.0001)	0.123
Maths	952 (11)	68.192 (68.140; 68.245)	995 (20)	69.937 (69.888; 69.986)	1877 (913; 964)	-0.146 (-0.235; -0.057)	0.016
Self-esteem	725 (365)	11.446 (11.441; 11.451)	733 (427)	11.416 (11.411; 11.421)	1353 (667; 686)	-0.004 (-0.106; 0.099)	0.949
Goals and aspirations	728 (362)	8.398 (8.394; 8.403)	728 (432)	8.129 (8.124; 8.134)	1348 (671; 677)	0.156 (0.053; 0.259)	0.012
Family connection	712 (378)	17.857 (17.850; 17.863)	714 (446)	17.682 (17.675; 17.689)	1272 (617; 655)	0.084 (-0.020; 0.188)	0.184
School connection	712 (378)	17.360 (17.351; 17.368)	704 (456)	17.297 (17.289; 17.305)	1272 (635; 637)	0.059 (-0.045; 0.163)	0.354
Subgroup effects of intervention—pupils in the AfA target group (Model 2.2)							
Reading	594 (63)	25.981 (25.947; 26.016)	653 (40)	27.435 (27.404; 27.466)	1224 (579; 645)	-0.151 (-0.262; -0.040)	0.030
Maths	600 (57)	58.462 (58.376; 58.548)	652 (41)	62.130 (62.050; 62.210)	1229 (585; 644)	-0.147 (-0.258; -0.036)	0.041
Self-esteem	479 (178)	11.248 (11.241; 11.256)	454 (239)	11.139 (11.131; 11.147)	869 (447; 422)	0.053 (-0.075; 0.181)	0.464
Goals and aspirations	478 (179)	8.182 (8.175; 8.189)	448 (245)	8.031 (8.023; 8.039)	859 (447; 412)	0.098 (-0.031; 0.227)	0.181
Family connection	472 (185)	17.756 (17.746; 17.767)	431 (262)	17.348 (17.335; 17.361)	812 (420; 392)	0.118 (-0.013; 0.249)	0.091
School connection	471 (186)	17.335 (17.323; 17.347)	438 (255)	17.126 (17.112; 17.139)	821 (430; 391)	0.082 (-0.048; 0.212)	0.236
Absence	634 (23)	1.90% (1.89%; 1.91%)	680 (13)	1.55% (1.54%; 1.56%)	1313 (633; 680)	IRR= 1.053 (0.802; 1.383)	0.709

The values in this table have been calculated based on complete cases as per EEF guidance. *p*-values in this table correspond to the co-efficient in a given model, whereas CIs correspond to the standardised intervention effect size.

Primary outcome—reading

Our complete case ITT analysis revealed a negative effect of AfA on pupils' reading scores (ES: -0.096; CI: -0.146 to -0.046). This effect was not sensitive to the inclusion of additional covariates (%FSM, %SEN, %RWM4+, and usual practice indicator at the school level; gender and SEN at the pupil level) or multiple imputation of missing data. In summary, our analyses indicated that business as usual was superior to AfA in improving pupils' reading scores.

Secondary outcome—mathematics

Our complete case ITT analysis revealed a negative effect of AfA on pupils' maths scores (ES: -0.128; CI: -0.178 to -0.078). This effect was not sensitive to the inclusion of additional covariates or multiple imputation of missing data. In summary, our analyses indicated that business as usual was superior to AfA in improving pupils' maths scores.

Secondary outcomes—resilience-related measures

Our complete case analysis revealed no effects of AfA on pupils' self-esteem, goals and aspirations, family connection, or school connection. These null effects were not sensitive to the inclusion of additional covariates or multiple imputation of missing data. In summary, our analyses indicated AfA had no impact on pupils' resilience-related outcomes.

Secondary outcome—attendance

Our complete case analysis revealed no effects of AfA on target pupils' attendance. These null effects were not sensitive to the inclusion of additional covariates (note that multiple imputation was not undertaken as a sensitivity analysis as missing data did not exceed 5%).

In summary, our analyses indicated AfA had no impact on target pupils' attendance.

Subgroup analyses

FSM

Our complete case subgroup analyses revealed a negative effect of AfA on FSM-eligible pupils' maths scores (ES: -0.146; CI: -0.235 to 0.057) and a positive effect on their goals and aspirations (ES: 0.156; CI: 0.053 to 0.259) but no apparent impact on their reading, self-esteem, family connection, or school connection. These findings were insensitive to changes in our modelling parameters except that a negative effect on pupils' reading scores became evident when missing data was multiply imputed. This is likely due to increased test sensitivity resulting from a greater volume of data being available when using MI.

In summary, our analyses indicated that business as usual was superior to AfA in improving FSM-eligible pupils' maths scores; under certain modelling specifications (for example, MI), this was also evident for their reading scores. By contrast, AfA was superior to business as usual in improving the goals and aspirations of FSM-eligible pupils.

AfA target group

Our complete case subgroup analyses revealed a negative effect of AfA on target pupils' reading scores (ES: -0.151; CI: -0.262 to -0.040) and maths scores (ES: -0.147; CI: -0.258 to -0.036), but no apparent impact on their self-esteem, goals and aspirations, family connection, or school connection. These findings were insensitive to changes in our modelling parameters. In summary, our analyses indicated that business as usual was superior to AfA in improving target pupils' reading and maths scores.

Comparison of main report (Cohort 1/Y5) and addendum report (Cohort 2/Y4) impact findings

Table 11 provides a simple summary of our complete case analysis findings relating to the impact of AfA across the two trial cohorts. Inspection of said table reveals that ITT findings are identical across the two cohorts, with the exception of a positive impact of AfA on school connectedness for Cohort 1 not being replicated in Cohort 2. Findings pertaining to the FSM subgroup are very similar with the following exceptions: (i) the negative impact of AfA on reading scores in FSM Cohort 1 was not replicated in FSM Cohort 2 and (ii) the positive impact of AfA on goals and aspirations of FSM Cohort 2 was not evident for FSM Cohort 1. Finally, findings pertaining to the AfA target group are identical across the two trial cohorts.

Table 11: Simple comparison of complete case impact analyses for AfA trial cohorts

	Main report (Cohort 1, Y5) interpretation	Effect size	Addendum report (Cohort 2, Y4) interpretation	Effect size
Reading ITT	Negative	-0.119	Negative	-0.096
Maths ITT	Negative	-0.114	Negative	-0.128
Self-esteem ITT	Null	-0.005	Null	0.002
Goals and aspirations ITT	Null	0.045	Null	0.063
Family connection ITT	Null	0.061	Null	0.061
School connection ITT	Positive	0.152	Null	0.088
Reading FSM subgroup	Negative	-0.123	Null*	-0.089
Maths FSM subgroup	Negative	-0.169	Negative	-0.146
Self-esteem FSM subgroup	Null	-0.037	Null	-0.004
Goals and aspirations FSM subgroup	Null	0.006	Positive	0.156
Family connection FSM subgroup	Null	0.039	Null	0.084
School connection FSM subgroup	Null	0.125	Null	0.059
Reading AfA target subgroup	Negative	-0.157	Negative	-0.151
Maths AfA target subgroup	Negative	-0.185	Negative	-0.147
Self-esteem AfA target subgroup	Null	0.029	Null	0.053
Goals and aspirations AfA target subgroup	Null	0.016	Null	0.098
Family connection AfA target subgroup	Null	-0.029	Null	0.118
School connection AfA target subgroup	Null	0.152	Null	0.082
Attendance AfA target subgroup	Null	-0.005	Null	IRR 1.053

Note: 'Interpretation' columns are based on a combination of the p-value and coefficient direction.

*This null result became negative in the sensitivity analysis when missing data was multiply imputed.

Conclusion

Key conclusions

1. Children in the Achievement for All schools made two months less progress in reading, on average, compared to children in schools that did not receive the programme.
2. Target children in the Achievement for All schools (the lowest 20% of attainers or those deemed to be 'vulnerable to underachievement' as identified by their school) made two months less progress in reading, on average, compared to target children in schools that did not receive the programme. Children eligible for FSM in Achievement for All schools made one month less progress in reading, on average, compared to FSM-eligible children in schools that did not receive the programme.
3. All children, AfA target children, and FSM-eligible children in the Achievement for All schools made two months less progress in maths, on average, compared to equivalent children in schools that did not receive the programme.
4. The evaluation found that the programme did not improve pupils' self-esteem, goals and aspirations, perceptions of how supportive their families were, perceptions of how supportive their schools were, or target children's attendance. However, FSM-eligible children in Achievement for All schools were more likely to report higher levels of goals and aspirations.
5. The findings noted above are largely consistent with those documented in the main AfA trial report.

Interpretation

The purpose of this addendum report was to examine the impact of AfA on academic and resilience-related outcomes among a cohort of pupils exposed to the intervention for a full six school terms. Given the general pattern of negative and null findings in our main report (Humphrey et al., 2020), which were based on a cohort of pupils who only received five terms of exposure, the analyses reported herein provide an important opportunity to clarify the efficacy of the AfA programme when it is implemented for the full length of time recommended by the developer.

The findings reported here are in fact remarkably similar to those documented in our main report. Collectively, they provide robust evidence that AfA is inferior to the usual practice of schools. In both trial cohorts, usual practice was demonstrably superior to the intervention in terms of improving reading and maths scores of all children and the AfA target subgroup, and the maths scores of the FSM subgroup. The associated effect sizes equate to between two and three months less progress in these outcomes among children in AfA schools. The findings for Cohort 2/Y4—our focus in this report—were completely insensitive to changes in our modelling parameters, with one key exception: when missing data was accounted for via MI, a negative effect of AfA on FSM-eligible children's reading scores was evident. This is likely due to increased test sensitivity resulting from a greater volume of data being available when using MI.

In our main report, we discussed the fact that these findings place AfA on the first percentile in the distribution of intervention effect sizes for academic outcomes of universal, school-based interventions (Tanner-Smith, Durlak and Marx, 2018) and noted the programme's failure to address policy relevant gaps for vulnerable subgroups such as those eligible for FSM (Hill et al., 2008). Our findings for Cohort 2/Y4 reinforce these messages and, importantly, they enable us to rule out insufficient exposure as an explanation. It always seemed likely that this would be the case—it is difficult to imagine a scenario in which five terms of exposure produces negative effects but a single additional term somehow 'turns things around' and produces meaningful positive impacts—but it was nonetheless important to provide a full and fair test of the intervention.

Three key differences between the findings of our main report and this addendum report are worthy of brief discussion:

- a positive impact of AfA on school connectedness for Cohort 1 was not replicated in Cohort 2;
- a positive impact of AfA on goals and aspirations of FSM Cohort 2 that was not evident for FSM Cohort 1; and
- a negative impact of AfA on reading scores in FSM Cohort 1 was not replicated in FSM Cohort 2.

It is difficult to provide a clear explanation for the discrepancies in the two resilience-related outcomes (school connectedness and goals and aspirations) because the timing of assessments for these outcomes was in fact the same for both cohorts. Hence one would struggle to make the argument that the extra term of exposure to AfA experienced

by FSM Cohort 2 triggered the effect observed here in relation to their goals and aspirations since this post-test data was collected after only five terms. Ultimately, speculation about these mixed effects is probably moot anyway as one would not recommend an intervention that promotes school connectedness or goals and aspirations if this were at the expense of academic attainment (as is the case here).

By contrast, the post-test reading scores for FSM Cohort 2 did follow six terms of exposure to AfA. Here, we believe that the null effect results from marginally lower test sensitivity in the complete case analysis for Cohort 2 (compared to Cohort 1). The emergence of a negative effect in our sensitivity analysis using MI of missing data supports this proposition: under these modelling conditions, we replicate the effect seen in the main report.

Strengths and limitations

This study has numerous strengths, reflecting those of the main report, that increase confidence in the security of our findings:

- use of a cluster-randomised design with appropriate analysis that took account of the hierarchical and clustered nature of the dataset;
- a very large and well-powered trial, with an MDSE of 0.137 at the point of randomisation for Cohort 2/Y4;
- in terms of generalizability, the 134 trial schools spanned 78 (23%) of the 343 LAs across England;
- attrition was 0% at the school level and 7.77% at the pupil level for the primary outcome (reading); furthermore, there was no evidence of differential attrition by trial arm in our analysis of missing data;
- the use of a randomised design (with the allocation sequence conducted independently of the research team) meant that, in expectation, we would be free from confounders; in practice, balance on observables was indeed very good with negligible differences between pupil-level outcomes at baseline;
- our primary outcome (reading scores derived from national assessments at the end of KS2) has demonstrable reliability, validity, utility, and acceptability in relation to our target population; those responsible for grading these assessments were blind to trial group allocation (though of course it was not possible to achieve blinding in the administration of the tests, however, this is not expected to have introduced additional bias); and
- the analysis reported herein reflects the impact of AfA following the developer-recommended six terms of exposure to the intervention addressing a limitation of the main report.

Nonetheless, there are limitations to be borne in mind:

- AfA is inherently non-manualised and flexible as a programme. This meant that we were not able to apply complier average causal effect estimation (CACE) or related instrumental variable approaches to robustly account for the role of implementation variability in intervention outcomes.
- While the academic attainment and attendance data for Cohort 2/Y4 reflect the full, developer-recommended six terms of AfA exposure, the resilience-related secondary outcomes for Cohort 2/Y4 mirror the timings of Cohort 1/Y5, that is, five terms.

In conclusion, there is now compelling evidence that demonstrates the negative impact of AfA on the academic outcomes of all pupils, those eligible for FSM, and those identified as 'vulnerable to underachievement' (the priority group—'AfA target' children). The strength of confidence in this statement is derived from very similar findings across the two trial cohorts in this project alongside the highest possible EEF security rating being awarded to the main report. Thus, where there is independent, robust evidence available, it indicates that usual practice is superior to AfA in terms of improving academic outcomes.

References

- Barlow, A. et al. (2015) 'Evaluation of the implementation and impact of an integrated prevention model on the academic progress of students with disabilities', *Research in Developmental Disabilities*, 36. doi: 10.1016/j.ridd.2014.10.029.
- Barlow, A., Humphrey, N., Lendrum, A., Wigelsworth, M. and Squires, G. (2015) 'Evaluation of the Implementation and Impact of an Integrated Prevention Model on the Academic Progress of Students With Disabilities', *Research in Developmental Disabilities*, 36. DOI: 10.1016/j.ridd.2014.10.029
- Bloom, H., Richburg-Hayes, L. and Black, A. R. (2007) 'Using Covariates to Improve Precision for Studies that Randomise Schools to Evaluate Educational Interventions', *Educational Evaluation and Policy Analysis*, 29, pp. 30–59.
- Bloom, H., Richburg-Hayes, L. and Black, A.R. (2007) Using covariates to improve precision for studies that randomise schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, 29, 30-59.
- Churches, R. (2016) 'Closing the Gap: Test and Learn', London: Department for Education.
- Churches, R. (2016) Closing the gap: test and learn. London: DFE.
- Education Endowment Foundation. (2013). Pre-testing in EEF evaluations. London: EEF. Available at: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/Pre-testing_paper.pdf
- EEF (2013) 'Pre-Testing in EEF Evaluations', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/Pre-testing_paper.pdf
- Hedges, L. V. (2007) 'Effect Sizes in Cluster-Randomized Designs', *Journal of Educational and Behavioral Statistics*. doi: 10.3102/1076998606298043.
- Hedges, L. V. (2007) 'Effect Sizes in Cluster-Randomized Designs', *Journal of Educational and Behavioral Statistics*. DOI: 10.3102/1076998606298043
- Hill, C. et al. (2008) 'Empirical benchmarks for interpreting effect sizes in research', *Child Development ...*, 2(3), pp. 172–177.
- Hill, C. J., Bloom, H. S., Black, A. R. and Lipsey, M. W. (2008) 'Empirical Benchmarks for Interpreting Effect Sizes in Research', *Child Development Perspectives*, 2 (3), pp. 172–177.
- Humphrey, N. and Squires, G. (2010) 'Achievement for All Evaluation: Interim Report (May 2010)', Department for Education, Research Report DFE-RR028.
- Humphrey, N. and Squires, G. (2010) Achievement for All National Evaluation: Interim Report. Nottingham.
- Humphrey, N. and Squires, G. (2011a) 'Achievement for All National Evaluation: Final Report', Department for Education, Research Report DFE-RR176.
- Humphrey, N. and Squires, G. (2011a) Achievement for All National Evaluation: Final Report, Education. Nottingham.
- Humphrey, N. and Squires, G. (2011b) 'Achievement for All national Evaluation: Second Interim Report', Nottingham. Department for Education.
- Humphrey, N. and Squires, G. (2011b) Achievement for All national Evaluation: Second Interim Report. Nottingham.
- Humphrey, N. et al (2020): Achievement for All: Evaluation. London: EEF.
- Humphrey, N. et al. (2013) 'Achievement for All : Improving psychosocial outcomes for students with special educational needs and disabilities', *Research in developmental disabilities*, 34, pp. 1210–1225.
- Humphrey, N., Lendrum, A., Barlow, A., Wiglesworth, M. and Squires, G. (2013) 'Achievement for All : Improving Psychosocial Outcomes for Students With Special Educational Needs and Disabilities', *Research in Developmental Disabilities*, 34, pp. 1210–1225.

Humphrey, N., Squires, G., Choudry, S., Byrne, E., Demkowicz, O., Troncoso, P. and Wo, L. (2020) 'Achievement for All: Evaluation Report', London: Education Endowment Foundation. [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Achievement_for_All_\(final\).pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Achievement_for_All_(final).pdf)

Lereya, S. T. et al. (2016) 'The student resilience survey: psychometric validation and associations with mental health', *Child and Adolescent Psychiatry and Mental Health*. BioMed Central, 10(1), p. 44. doi: 10.1186/s13034-016-0132-5.

Lereya, S. T., Humphrey, N., Patalay, P., Wolpert, M., Böhnke, J. R., Macdougall, A. and Deighton, J. (2016) 'The Student Resilience Survey: Psychometric Validation and Associations With Mental Health', *Child and Adolescent Psychiatry and Mental Health*, 10, article no. 44. DOI: 10.1186/s13034-016-0132-5

Sun, J. and Stewart, D. (2007) 'Development of population-based resilience measures in the primary school setting', *Health Education*, 107(6), pp. 575–599. doi: 10.1108/09654280710827957.

Sun, J. and Stewart, D. (2007) 'Development of Population-Based Resilience Measures in the Primary School Setting', *Health Education*, 107 (6), pp. 575–599. DOI: 10.1108/09654280710827957

Tanner-Smith, E. E., Durlak, J. A. and Marx, R. A. (2018) 'Empirically Based Mean Effect Size Distributions for Universal Prevention Programs Targeting School-Aged Youth: A Review of Meta-Analyses', *Prevention Science*. *Prevention Science*, 19(8), pp. 1091–1101. doi: 10.1007/s11121-018-0942-1.

Tanner-Smith, E. E., Durlak, J. A. and Marx, R. A. (2018) 'Empirically Based Mean Effect Size Distributions for Universal Prevention Programs Targeting School-Aged Youth: A Review of Meta-Analyses', *Prevention Science*, 19 (8), pp. 1091–1101. DOI: 10.1007/s11121-018-0942-1

Wigelsworth, M. et al (2018) *FRIENDS for Life: Evaluation report and executive summary*. London: EEF.

Wigelsworth, M., Squires, G., Birchinall, L., Kalambouka, A., Lendrum, A., Black, L., Troncoso, P., Santos, J., Ashworth E. and Britteon, P. (2018) 'FRIENDS for Life: Evaluation Report and Executive Summary', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/FRIENDS.pdf

Appendix 1: MLM ITT and subgroup analyses

Table A1: Reading – Whole phase 2 cohort – complete case

Level	Empty model			Model 1.1			Model 1.2		
	Coef.	SE	p	Coef.	SE	p	Coef.	SE	p
Pupil Variance	90.726	1.644	--	50.663	0.930	--	49.493	0.908	--
Pre-test (baseline)				1.814	0.026	<0.001	1.673	0.029	<0.001
Gender (if male)							-0.996	0.185	<0.001
SEND (if eligible)							-2.922	0.306	<0.001
School Variance	6.096	1.022	--	3.829	0.634	--	3.155	0.540	--
Trial group (if AfA)				-0.947	0.394	0.016	-1.087	0.367	0.003
Rand. variables (low as reference):	%FSM-Medium						-0.298	0.478	0.533
	%FSM-High						-1.307	0.562	0.020
	%SEND-Medium						0.217	0.460	0.637
	%SEND-High						1.189	0.550	0.031
	%RWM+4-Medium						0.481	0.459	0.295
%RWM+4-High						1.252	0.482	0.009	
Usual practice score							0.081	0.057	0.157
Loglikelihood	-22927.869			-20634.03			-20531.2		
VPC	0.063			0.070			0.060		
n	6218			6074			6067		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A2: Reading – FSM and AFA subgroups – complete case

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	59.403	2.023	--	68.525	2.933	--
Pre-test (baseline)	1.785	0.052	<0.001	1.630	0.073	<0.001
School Variance	5.769	1.426	--	8.952	2.187	--
Trial group (if AfA)	-0.906	0.587	0.123	-1.583	0.729	0.030
Loglikelihood	-6485.9325			-4373.5898		
VPC	0.089			0.116		
n	1859			1224		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A3: Mathematics – Whole phase 2 cohort – complete case

Level	Empty model			Model 1.1			Model 1.2		
	Intercept (SE) = 76.439 (0.677)			Intercept (SE) = -4.621 (1.358)			Intercept (SE) = -1.821 (2.663)		
	Coef.	SE	p	Coef.	SE	p	Coef.	SE	p
Pupil Variance	522.930	9.479	--	267.379	4.905	--	260.093	4.774	--
Pre-test (baseline)				5.024	0.067	<0.001	4.639	0.072	<0.001
Gender (if male)							1.667	0.424	<0.001
SEND (if eligible)							-8.810	0.698	<0.001
School Variance	47.433	7.623	--	38.201	5.697	--	34.557	5.203	--
Trial group (if AfA)				-3.040	1.169	0.009	-3.413	1.121	0.002
Rand. variables (low as reference):	%FSM-Medium						1.024	1.460	0.483
	%FSM-High						-0.726	1.710	0.671
	%SEND-Medium						-0.254	1.406	0.857
	%SEND-High						1.122	1.671	0.502
	%RWM+4-Medium						0.422	1.404	0.764
%RWM+4-High						1.405	1.472	0.340	
Usual practice score							0.318	0.176	0.070
Loglikelihood	-28420.506			-25753.434			-25635.91		
VPC	0.083			0.125			0.117		
n	6225			6082			6075		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A4: Maths – FSM and AFA subgroups – complete case

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Intercept (SE) = -9.152 (2.341)			Intercept (SE) = -11.776 (3.040)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	328.424	11.130	--	374.058	15.967	--
Pre-test (baseline)	5.096	0.136	<0.001	5.156	0.194	<0.001
School Variance	43.078	9.209	--	71.824	14.814	--
Trial group (if AfA)	-3.659	1.512	0.016	-3.886	1.905	0.041
Loglikelihood	-8133.66			-5448.4843		
VPC	0.116			0.161		
n	1870			1229		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A5: Self-esteem – Whole phase 2 cohort – complete case

Level	Empty model			Model 1.1			Model 1.2		
	Coef.	SE	p	Coef.	SE	p	Coef.	SE	p
Pupil Variance	3.201	0.067	--	3.013	0.065	--	2.992	0.065	--
Pre-test (baseline)				0.235	0.014	<0.001	0.220	0.014	<0.001
Gender (if male)							-0.033	0.053	0.531
SEND (if eligible)							-0.528	0.080	<0.001
School Variance	0.126	0.028	--	0.091	0.024	--	0.058	0.019	--
Trial group (if AfA)				0.003	0.079	0.969	-0.025	0.072	0.724
Rand. variables (low as reference):	%FSM-Medium						-0.309	0.093	0.001
	%FSM-High						-0.155	0.111	0.161
	%SEND-Medium						-0.170	0.091	0.061
	%SEND-High						-0.084	0.108	0.436
	%RWM+4-Medium						0.077	0.091	0.395
%RWM+4-High						0.063	0.095	0.508	
Usual practice score							0.009	0.012	0.432
Loglikelihood	-9507.395			-8796.4247			-8637.9438		
VPC	0.039			0.029			0.019		
n	4726			4443			4376		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A6: Self-esteem – FSM and AFA subgroups – complete case

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	3.343	0.132	--	3.241	0.164	--
Pre-test (baseline)	0.180	0.025	<0.001	0.062	0.031	0.047
School Variance	0.021	0.035	--	0.061	0.062	--
Trial group (if AfA)	-0.007	0.105	0.949	0.098	0.133	0.464
Loglikelihood	-2740.2321			-1751.556		
VPC	0.006			0.019		
n	1353			869		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A7: Goals and aspirations – Whole phase 2 cohort – complete case

Level	Empty model			Model 1.1			Model 1.2		
	Coef.	SE	p	Coef.	SE	p	Coef.	SE	p
Pupil Variance	2.820	0.059	--	2.697	0.058	--	2.691	0.058	--
Pre-test (baseline)				0.196	0.014	<0.001	0.189	0.014	<0.001
Gender (if male)							-0.068	0.051	0.178
SEND (if eligible)							-0.370	0.076	<0.001
School Variance	0.108	0.024	--	0.077	0.020	--	0.068	0.019	--
Trial group (if AfA)				0.108	0.074	0.144	0.042	0.093	0.652
Rand. variables (low as reference):	%FSM-Medium						0.042	0.093	0.652
	%FSM-High						0.040	0.111	0.722
	%SEND-Medium						-0.009	0.091	0.923
	%SEND-High						-0.001	0.109	0.991
	%RWM+4-Medium						0.070	0.092	0.444
%RWM+4-High						0.014	0.095	0.886	
Usual practice score							0.026	0.012	0.030
Loglikelihood	-9216.5498			-8547.3167			-8413.8979		
VPC	0.037			0.028			0.025		
n	4731			4442			4377		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A8: Goals and aspirations – FSM and AFA subgroups – complete case

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	2.985	0.119	--	2.911	0.150	--
Pre-test (baseline)	0.195	0.026	<0.001	0.144	0.031	<0.001
School Variance	0.071	0.041	--	0.069	0.064	--
Trial group (if AfA)	0.278	0.111	0.012	0.173	0.129	0.181
Loglikelihood	-2663.1207			-1686.8954		
VPC	0.023			0.023		
n	1348			859		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A9: Family connection – Whole phase 2 cohort – complete case

Level	Empty model			Model 1.1			Model 1.2		
	Coef.	SE	p	Coef.	SE	p	Coef.	SE	p
Pupil Variance	4.929	0.104	--	4.498	0.100	--	4.390	0.098	--
Pre-test (baseline)				0.218	0.013	<0.001	0.196	0.013	<0.001
Gender (if male)							-0.367	0.067	<0.001
SEND (if eligible)							-0.596	0.101	<0.001
School Variance	0.194	0.043	--	0.155	0.037	-	0.148	0.037	--
Trial group (if AfA)				0.137	0.102	0.180	0.128	0.101	0.206
Rand. variables (low as reference):	%FSM-Medium						-0.107	0.131	0.412
	%FSM-High						0.057	0.155	0.715
	%SEND-Medium						0.004	0.128	0.972
	%SEND-High						-0.155	0.152	0.308
	%RWM+4-Medium						0.071	0.128	0.580
%RWM+4-High						0.021	0.133	0.876	
Usual practice score							-0.007	0.017	0.684
Loglikelihood	-10260.352			-9098.142			-8906.0563		
VPC	0.038			0.033			0.033		
n	4606			4171			4106		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A10: Family connection – FSM and AFA subgroups – complete case

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	5.272	0.216	--	6.296	0.330	--
Pre-test (baseline)	0.217	0.024	<0.001	0.225	0.029	<0.001
School Variance	0.158	0.081	--	0.309	0.183	0.091
Trial group (if AfA)	0.207	0.156	0.184	0.052	0.115	--
Loglikelihood	-2877.746			-1902.4387		
VPC	0.029			0.008		
n	1272			812		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A11: School connection – Whole phase 2 cohort – complete case

		Empty model Intercept (SE) = 17.271 (0.075)			Model 1.1 Intercept (SE) = 13.016 (0.298)			Model 1.2 Intercept (SE) = 13.134 (0.420)		
Level		Coef.	SE	p	Coef.	SE	p	Coef.	SE	p
Pupil Variance		8.232	0.173	--	7.703	0.170	--	7.613	0.169	--
	Pre-test (baseline)				0.238	0.016	<0.001	0.232	0.016	<0.001
	Gender (if male)							-0.431	0.087	<0.001
	SEND (if eligible)							-0.055	0.130	0.671
School Variance		0.393	0.083	--	0.328	0.075	--	0.285	0.072	--
	Trial group (if AfA)				0.259	0.142	0.068	0.268	0.137	0.050
Rand. variables (low as reference):	%FSM-Medium							-0.089	0.177	0.614
	%FSM-High							0.388	0.211	0.066
	%SEND-Medium							-0.159	0.173	0.357
	%SEND-High							-0.224	0.205	0.277
	%RWM+4-Medium							-0.040	0.174	0.820
	%RWM+4-High							-0.044	0.180	0.809
Usual practice score								0.027	0.023	0.228
Loglikelihood		-11515.411			-10336.992			-10154.502		
VPC		0.046			0.041			0.036		
n		4633			4216			4153		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A12: School connection – FSM and AFA subgroups – complete case

		FSM subgroup (model 2.1) Intercept (SE) = 13.300 (0.519)			AFA subgroup (model 2.2) Intercept (SE) = 13.619 (0.596)		
Level		Coef.	SE	p	Coef.	SE	p
Pupil Variance		8.152	0.334		8.210	0.426	
	Pre-test (baseline)	0.228	0.029	<0.001	0.205	0.034	<0.001
School Variance		0.197	0.114		0.060	0.141	
	Trial group (if AfA)	0.175	0.188	0.354	0.245	0.207	0.236
Loglikelihood		-3152.3002			-2032.0618		
VPC		0.024			0.007		
n		1272			821		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A13: Absence – AfA subgroup phase 2 cohort – complete case

		Empty model			Model 1.1			Model 1.2		
		Intercept (SE) = -4.211 (0.076)			Intercept (SE) = -4.727 (0.104)			Intercept (SE) = -4.747 (0.294)		
Level		Coef.	SE	p	Coef.	SE	p	Coef.	SE	p
Pupil level										
	Pre-test (baseline)				0.063	0.007	0.000	0.062	0.007	0.000
	Gender (if male)							-0.055	0.111	0.621
	SEND (if eligible)							-0.004	0.114	0.970
	Overdispersion	1.279	0.051	--	1.148	0.053	--	1.148	0.054	
School Variance		0.260	0.081	--	0.215	0.074	--	0.176	0.070	
	Trial group (if AfA)				0.052	0.139	0.709	0.061	0.135	0.654
	Rand. variables (low as reference):	%FSM-Medium						0.235	0.176	0.182
		%FSM-High						0.296	0.212	0.164
		%SEND-Medium						0.162	0.167	0.334
		%SEND-High						-0.169	0.208	0.418
		%RWM+4-Medium						-0.201	0.165	0.224
	%RWM+4-High							-0.249	0.180	0.166
	Usual practice score							0.003	0.021	0.873
	Loglikelihood	-3317.9284			-3253.5404			-3238.0857		
	VPC	0.005			0.002			0.002		
	n	1314			1313			1308		

Note. Models were computed in Stata 16.1 using ML estimation.

Table A24: Reading – Whole phase 2 cohort – Multilevel multiple imputation

Level	Model 1.1			Model 1.2		
	Intercept (SE) = 3.331 (0.504)			Intercept (SE) = 5.432 (0.915)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	50.997	--	--	50.025	--	--
Pre-test (baseline)	1.820	0.025	<0.001	1.674	0.028	<0.001
Gender (if male)				-0.987	0.185	<0.001
SEND (if eligible)				-2.718	0.311	<0.001
School Variance	3.805	--	--	3.048	--	--
Trial group (if AfA)	-1.020	0.391	0.009	-1.133	0.361	0.002
Rand. variables (low as reference):	%FSM-Medium			-0.268	0.473	0.570
	%FSM-High			-1.321	0.558	0.018
	%SEND-Medium			0.198	0.455	0.664
	%SEND-High			1.113	0.545	0.041
	%RWM+4-Medium			0.466	0.457	0.307
	%RWM+4-High			1.310	0.476	0.006
Usual practice score				0.082	0.057	0.147
VPC		0.069			0.057	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package

Table A15: Reading – FSM and AFA subgroups – Multilevel multiple imputation

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Intercept (SE) = 4.434 (0.519)			Intercept (SE) = 5.601 (0.575)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	50.529	--	--	44.827	--	--
Pre-test (baseline)	1.785	0.026	<0.001	1.711	0.030	<0.001
FSM	-1.701	0.290	<0.001	--	--	--
FSM*Trial group (if AfA)	0.065	0.419	0.877	--	--	--
School Variance	3.561	--	--	3.290	--	--
Trial group (if AfA)	-0.995	0.405	0.014	-0.822	0.377	0.029
VPC		0.066			0.068	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package. The imputation model for the FSM subgroup differs from the SAP, as FSM eligibility is missing for some pupils. The alternative specification measures the effect of trial in the FSM subgroup as a cross-level interaction effect (FSM*Trial group).

Table A36: Maths – Whole phase 2 cohort – Multilevel multiple imputation

Level	Model 1.1			Model 1.2		
	Intercept (SE) = -4.605 (1.338)			Intercept (SE) = -1.600 (2.671)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	267.581	--	--	261.084	--	--
Pre-test (baseline)	5.026	0.065	<0.001	4.628	0.071	<0.001
Gender (if male)				1.611	0.424	<0.001
SEND (if eligible)				-8.392	0.701	<0.001
School Variance	38.298	--	--	34.832	--	--
Trial group (if AfA)	-3.154	1.173	0.007	-3.522	1.128	<0.001
Rand. variables (low as reference):	%FSM-Medium			1.024	1.467	0.485
	%FSM-High			-0.641	1.720	0.709
	%SEND-Medium			-0.244	1.410	0.863
	%SEND-High			1.060	1.681	0.528
	%RWM+4-Medium			0.289	1.414	0.838
	%RWM+4-High			1.492	1.477	0.312
Usual practice score				0.321	0.176	0.069
VPC		0.125			0.118	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package

Table A17: Maths – FSM and AfA subgroups – Multilevel multiple imputation

Level	FSM subgroup (model 2.1)			AfA subgroup (model 2.2)		
	Intercept (SE) = -1.092 (1.372)			Intercept (SE) = 4.879 (1.484)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	261.830	--	--	227.463	--	--
Pre-test (baseline)	4.901	0.065	<0.001	4.537	0.075	<0.001
FSM	-4.877	0.668	<0.001	--	--	--
FSM*Trial group (if AfA)	-1.394	0.972	0.152	--	--	--
School Variance	38.108	--	--	31.923	--	--
Trial group (if AfA)	-2.540	1.207	0.035	-2.547	1.098	0.020
VPC		0.127			0.123	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package. The imputation model for the FSM subgroup differs from the SAP, as FSM eligibility is missing for some pupils. The alternative specification measures the effect of trial in the FSM subgroup as a cross-level interaction effect (FSM*Trial group).

Table A48: Self-esteem – Whole phase 2 cohort – Multilevel multiple imputation

Level	Model 1.1			Model 1.2		
	Intercept (SE) = 8.986 (0.170)			Intercept (SE) = 9.414 (0.250)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	3.019	--	--	2.984	--	--
Pre-test (baseline)	0.236	0.014	<0.001	0.222	0.014	<0.001
Gender (if male)				-0.015	0.058	0.798
SEND (if eligible)				-0.541	0.075	<0.001
School Variance	0.101	--	--	0.069	--	--
Trial group (if AfA)	0.014	0.080	0.861	-0.018	0.074	0.808
Rand. variables (low as reference):	%FSM-Medium			-0.299	0.094	0.002
	%FSM-High			-0.145	0.109	0.185
	%SEND-Medium			-0.186	0.086	0.030
	%SEND-High			-0.110	0.106	0.303
	%RWM+4-Medium			0.070	0.094	0.454
%RWM+4-High				0.040	0.101	0.691
Usual practice score				0.006	0.012	0.647
VPC		0.033			0.023	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package

Table A19: Self-esteem – FSM and AFA subgroups – Multilevel multiple imputation

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Intercept (SE) = 9.176 (0.175)			Intercept (SE) = 9.857 (0.320)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	2.995	--	--	3.176	--	--
Pre-test (baseline)	0.229	0.014	<0.001	0.120	0.028	<0.001
FSM	-0.349	0.080	<0.001	--	--	--
FSM*Trial group (if AfA)	-0.046	0.109	0.677	--	--	--
School Variance	0.094	--	--	0.103	--	--
Trial group (if AfA)	0.037	0.086	0.669	0.066	0.130	0.611
VPC		0.030			0.031	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package. The imputation model for the FSM subgroup differs from the SAP, as FSM eligibility is missing for some pupils. The alternative specification measures the effect of trial in the FSM subgroup as a cross-level interaction effect (FSM*Trial group).

Table A20: Goals and aspirations – Whole phase 2 cohort – Multilevel multiple imputation

Level	Model 1.1			Model 1.2		
	Intercept (SE) = 6.717 (0.118)			Intercept (SE) = 6.653 (0.193)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	2.712	--	--	2.695	--	--
Pre-test (baseline)	0.197	0.014	<0.001	0.188	0.014	<0.001
Gender (if male)				-0.066	0.047	0.163
SEND (if eligible)				-0.348	0.071	<0.001
School Variance	0.091	--	--	0.081	--	--
Trial group (if AfA)	0.092	0.077	0.234	0.071	0.075	0.345
Rand. variables (low as reference):	%FSM-Medium			0.039	0.098	0.694
	%FSM-High			0.078	0.113	0.491
	%SEND-Medium			-0.050	0.094	0.595
	%SEND-High			-0.052	0.108	0.627
	%RWM+4-Medium			0.061	0.097	0.532
	%RWM+4-High			0.004	0.094	0.964
Usual practice score				0.021	0.012	0.070
VPC		0.032			0.029	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package

Table A21: Goals and aspirations – FSM and AFA subgroups – Multilevel multiple imputation

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Intercept (SE) = 6.777 (0.118)			Intercept (SE) = 6.798 (0.237)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	2.709	--	--	2.877	--	--
Pre-test (baseline)	0.196	0.014	<0.001	0.162	0.029	<0.001
FSM	-0.173	0.074	0.021	--	--	--
FSM*Trial group (if AfA)	0.129	0.101	0.202	--	--	--
School Variance	0.090	--	--	0.087	--	--
Trial group (if AfA)	0.053	0.084	0.526	0.120	0.120	0.317
VPC		0.032			0.029	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package. The imputation model for the FSM subgroup differs from the SAP, as FSM eligibility is missing for some pupils. The alternative specification measures the effect of trial in the FSM subgroup as a cross-level interaction effect (FSM*Trial group).

Table A22: Family connection – Whole phase 2 cohort – Multilevel multiple imputation

Level	Model 1.1			Model 1.2		
	Intercept (SE) = 14.009 (0.258)			Intercept (SE) = 14.768 (0.347)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	4.619	--	--	4.529	--	--
Pre-test (baseline)	0.226	0.014	<0.001	0.206	0.014	<0.001
Gender (if male)				-0.383	0.067	<0.001
SEND (if eligible)				-0.587	0.104	<0.001
School Variance	0.150	--	--	0.137	--	--
Trial group (if AfA)	0.162	0.103	0.117	0.140	0.101	0.169
Rand. variables (low as reference):	%FSM-Medium			-0.136	0.127	0.285
	%FSM-High			0.019	0.152	0.903
	%SEND-Medium			0.019	0.121	0.873
	%SEND-High			-0.127	0.140	0.364
	%RWM+4-Medium			-0.003	0.129	0.981
	%RWM+4-High			-0.040	0.124	0.749
Usual practice score				-0.004	0.017	0.811
VPC		0.031			0.029	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package

Table A23: Family connection – FSM and AFA subgroups – Multilevel multiple imputation

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Intercept (SE) = 14.169 (0.255)			Intercept (SE) = 13.652 (0.446)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	4.605	--	--	5.769	--	--
Pre-test (baseline)	0.222	0.014	<0.001	0.225	0.026	<0.001
FSM	-0.301	0.100	0.003	--	--	--
FSM*Trial group (if AfA)	0.098	0.137	0.477	--	--	--
School Variance	0.149	--	--	0.109	--	--
Trial group (if AfA)	0.136	0.111	0.221	0.304	0.161	0.060
VPC		0.031			0.019	

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package. The imputation model for the FSM subgroup differs from the SAP, as FSM eligibility is missing for some pupils. The alternative specification measures the effect of trial in the FSM subgroup as a cross-level interaction effect (FSM*Trial group).

Table A24: School connection – Whole phase 2 cohort – Multilevel multiple imputation

Level	Model 1.1			Model 1.2		
	Intercept (SE) = 13.008 (0.285)			Intercept (SE) = 13.264 (0.395)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	7.893	--	--	7.849	--	--
Pre-test (baseline)	0.239	0.016	<0.001	0.232	0.016	<0.001
Gender (if male)				-0.421	0.087	0.519
SEND (if eligible)				0.025	0.119	0.835
School Variance	0.298	--	--	0.244	--	--
Trial group (if AfA)	0.236	0.135	0.082	0.229	0.129	0.077
Rand. variables (low as reference):	%FSM-Medium			-0.050	0.170	0.768
	%FSM-High			0.443	0.201	0.028
	%SEND-Medium			-0.181	0.160	0.258
	%SEND-High			-0.327	0.205	0.112
	%RWM+4-Medium			-0.110	0.160	0.492
	%RWM+4-High			-0.105	0.167	0.529
Usual practice score				0.020	0.022	0.363
VPC	0.036			0.030		

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package

Table A25: School connection – FSM and AFA subgroups – Multilevel multiple imputation

Level	FSM subgroup (model 2.1)			AFA subgroup (model 2.2)		
	Intercept (SE) = 12.954 (0.285)			Intercept (SE) = 13.446 (0.586)		
	Coef.	SE	p	Coef.	SE	p
Pupil Variance	7.890	--	--	8.203	--	--
Pre-test (baseline)	0.239	0.016	<0.001	0.218	0.034	<0.001
FSM	0.167	0.116	0.152	--	--	--
FSM*Trial group (if AfA)	-0.068	0.177	0.701	--	--	--
School Variance	0.293	--	--	0.187	--	--
Trial group (if AfA)	0.254	0.144	0.078	0.217	0.197	0.270
VPC	0.036			0.022		

Note. Estimates are based on the multilevel models implemented in the *lme4* R package over 20 multiply-imputed datasets pooled with the *mitml* R package. Imputation model was implemented in the *jomo* R package. The imputation model for the FSM subgroup differs from the SAP, as FSM eligibility is missing for some pupils. The alternative specification measures the effect of trial in the FSM subgroup as a cross-level interaction effect (FSM*Trial group).

This work was produced using statistical data from the ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit <https://nationalarchives.gov.uk/doc/open-government-licence/version/3> or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.



Education
Endowment
Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

<https://educationendowmentfoundation.org.uk>

 [@EducEndowFoundn](https://twitter.com/EducEndowFoundn)

 Facebook.com/EducEndowFoundn