Achievement for All

Evaluation Report

May 2020

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

The University of Manchester

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.

This work was produced using statistical data from 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.

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....3

Executive summary....4

Introduction....6

Methods....13

Impact evaluation....24

Implementation and process evaluation....36

Conclusion....60

References....66

Appendix A: Schematic of the AfA theory of change....69

Appendix B: Interview schedules used in case study schools....70

Appendix C: Memorandum of Agreement....74

Appendix D: Information sheet for parents/carers....79

Appendix E: General Data Protection Regulation notice....84

Appendix F: Student Resilience Survey (SRS) for the AfA trial....86

Appendix G: The AfA Teacher Implementation Survey....87

Appendix H: Histograms for the primary pre-test measures....100

Appendix I: MLM ITT and subgroup analyses....101

Appendix J: Multi-level negative binomial regression models for the attendance data127

Appendix K: The Usual Practice Survey....128

Appendix L: UPS data baseline comparison....134

Appendix M: UPS sub-category descriptives and ANOVAs....135

Appendix N: EEF cost rating....136

Appendix O: Security classification of trial findings....137

About the evaluator

The project was independently evaluated by a team from the Manchester Institute of Education, University of Manchester:

Neil Humphrey, Garry Squires, Sophina Choudry, Elizabeth Byrne, Ola Demkowicz, Patricio Troncoso, Lawrence Wo

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**

Executive summary

The project

Achievement for All (AfA) is a whole-school improvement programme that aims to improve the academic and social outcomes of pupils, developed by the national charity AfA 3As.[1] In this trial, the programme primarily aimed to improve Key Stage 2 reading attainment while the trial also examined its impact on KS2 mathematics, pupil attendance, and pupil resilience-related outcomes. The trial cohort comprised all children who began the trial in Years 4 and 5 (ages 8–10). A particular focus was placed on a target group of children selected by each school within these year groups. Schools were advised to select target children whose attainment placed them in the lowest 20%, while they could also include pupils whom they deemed to be vulnerable to underachievement.

The intervention ran for five terms. At the start of the programme, each school designated a member of staff to become an AfA champion. They then met with a trained AfA coach to assess the needs of the school and devised a bespoke action plan. This plan was then used to inform monthly coaching and training sessions delivered by the AfA coach to relevant members of school staff; schools also had access to AfA's online learning platform, The Bubble. AfA is a flexible programme that is expected to be tailored to the needs of each school. However, the training that each school receives draws from four key areas: 'leadership for inclusion', 'teaching and learning', 'wider outcomes and opportunities', and 'engaging with parents and carers'.

The project was a two-armed randomised controlled trial; 134 schools from across England participated, with 66 schools in the intervention group and 68 schools in the control group. The process evaluation included surveys, informal observations, and interviews, and a particular focus on eight case study schools. This report details the impact that the programme had on the first cohort of pupils who were in Year 5 at the outset of the trial in 2016/2017 and received the programme for five terms. The results from a second cohort, who began the trial in Year 4, will be examined in a future addendum report.

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. This result has a very high security rating.
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. This result has a very high security rating.
3.	All children and children eligible for free school meals (FSM) 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 target children made three months less progress in maths, on average, compared to target children in control schools. FSM children in Achievement for All schools also made two months less progress in reading compared to FSM children in schools that did not receive the intervention.
4.	The evaluation found that the programme did not improve pupils' self-esteem, goals and aspirations, perceptions of how supportive their families were, or the attendance of target children. However, children in Achievement for All schools were more likely to report that there was an adult in their school who cared about and supported them.
5.	The implementation of Achievement for All was not optimal and varied across schools. However, there was no evidence to suggest that this contributed to the negative findings. Some teachers identified significant resource demands which made implementing Achievement for All challenging.

EEF security rating

The findings related to the reading outcomes of all children and the reading outcomes of target children have a very high security rating. This trial was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools. The trial was a well-designed, two-armed, randomised controlled trial that was well powered. Relatively few pupils (8%) who started the trial were not included in the final analysis, while the pupils in Achievement for All schools were similar to those in comparison schools in terms of prior attainment.

[1] The '3 As' in AfA 3 As are Aspiration, Access and Achievement.

Additional findings

In this trial, Achievement for All resulted in negative impacts on pupils' academic outcomes. Children in the AfA schools made two months less progress in KS2 reading and maths compared to children in control schools—a finding replicated for children eligible for free school meals. Target children also made two months less progress in reading in addition to three months less progress in maths. The findings related to the reading outcomes of all children, and the reading outcomes of target children have a very high security rating. Given the size of the effects, the security of the primary outcome finding, and the consistency of the negative findings, these results are noteworthy. Even when accounting for the statistical uncertainty around the estimate, the possible impacts are consistently negative (ranging from one to four months less progress). The exception is the FSM reading result where there is a very small chance that the possible impact could have been null. Of particular importance is the impact that the programme had on target children and children eligible for free school meals. Both subgroups are considered vulnerable given the evidence that they are likely to experience disadvantageous outcomes during and following school. The evaluation found that the programme did not improve pupils' self-esteem, goals and aspirations, perceptions of how supportive their families were, or target children's attendance. However, children in AfA schools were more likely to report that there was an adult in their school who cared about and supported them.

There was variable implementation of the programme and schools failed to deliver key elements of the intervention. For instance, few schools delivered the expected number of mandatory 'structured conversations' to the parents of target children. It may also be the case that schools focused more on 'teaching and learning' and 'engaging with parents and carers' than they did on 'leadership for inclusion' or 'wider outcomes and opportunities'. There were also very few instances of staff using the online 'Bubble' resources in case study schools. However, 71% of schools did receive 20 or more coaching sessions (out of an intended 24); furthermore, the analysis of implementation suggested that higher levels of implementation were not associated with improved outcomes. More negative impacts may have been associated with higher fidelity implementation of the 'leadership for inclusion' strand of the programme.

One factor that may have contributed to the negative findings is the flexibility of the programme. While some case study schools commented that they found this beneficial, others reported that the flexibility made it difficult to grasp what the programme was. Several teachers within the case studies could not identify how the programme could lead to a direct impact on pupil learning. In addition, it may have been the case that other school priorities began to compete with the priorities set by the AfA action plan at the outset of the programme. School responses to Ofsted actions or other emerging issues may have overshadowed their action plans. In this context, and given these weaknesses, the negative findings may have, in part, been caused by the resources that schools expended on the programme at the expense of other activities. Champions and teachers in case studies commented that certain aspects of the programme came with a number of time and logistical demands, which may have had financial consequences. The implementation of termly, 'structured conversations', which are up to an hour long with the parents of all target children, was particularly resource intensive. On the basis of these highly secure findings, the EEF concludes that, in this trial, AfA did not improve pupils' academic outcomes and had a detrimental effect on learning.

Cost

The average cost of Achievement for All for one school was around £11.21 per pupil, per year, when averaged over three years. Schools are required to accommodate 12 half-day coaching visits each academic year (which may be attended by one or more teachers). Schools are also expected to conduct three 'structured conversations', which are up to an hour long, with the parents and carers of all target children each year.

Impact

Table 1: Summary of impact on primary outcome

Outcome/ group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	p value	EEF cost rating
Reading (whole group)	-0.12 (-0.17, -0.07)	-2	🔒🔒🔒🔒🔒	5,813	0.008	£ £ £ £ £
Reading (AfA target group)	-0.16 (-0.27, -0.05)	-2	🔒🔒🔒🔒🔒	1,231	0.026	£ £ £ £ £

Introduction

Background evidence

This report focuses on the impact of a school improvement programme, Achievement for All (AfA), on children's educational outcomes (such as reading) and resilience-related outcomes (for example, goals and aspirations). We examine outcomes at both the intent-to-treat and subgroup level. In relation to the latter, our focus is primarily on a target group that represents the lowest achieving 20% of children in a given school. This focus reflects the stated aims of the AfA programme in its current guise (see www.afaeducation.org) along with its history and development (AfA was originally designed as a programme to support children and young people with special educational needs and disabilities, SEND; Department for Children Schools and Families, 2009a). More broadly, it is recognized that responsiveness to intervention is unlikely to be uniform in a universal sample; as a result, there have been calls for researchers to engage more actively in pre-specified, defensible subgroup moderator analyses (Farrell, Henry and Bettencourt, 2013). Improving outcomes among this AfA target group is of the utmost importance. These children, who, based on analysis of our trial sample data, are likely to be *male* (1.3x compared to females), have *SEND* (9.8x more likely to have a statement of SEND/EHC plan, and 5x more likely to have SEND but without statement, compared to children with no SEND), be eligible for *FSM* (8.8x compared to those not eligible for FSM), and be *born in the final third of the school year* (1.25x compared to those born in the first third of the year), are among the most vulnerable in the education system (Squires et al., 2012). International research shows that those pupils in the AfA target groups are more likely to leave school early (European Agency for Special Needs and Inclusive Education, 2016, 2017a, 2017b), have lower lifelong earnings (Akbulut-Yuksel, 2017), and increased health and social care costs (OECD, 2010, 2012; Borgna and Struffolino, 2017). Addressing the needs of vulnerable groups has long been a priority for successive governments (see for example, Department for Children, Schools and Families, 2008). In international policy-making there has been a focus on inclusive education with aspirations expressed in the Salamanca agreement (UNESCO, 1994), Dakar agreement (UNESCO, 2000), and Incheon declaration (UNESCO, 2015a, 2015b). The focus of these declarations is increasingly underpinned by the rights of children and persons with disabilities to participate fully in all aspects of society without stigmatisation and marginalisation (UN General Assembly, 1989, 2006).

History, development, and evidence for the Achievement for All programme

Achievement for All began as a pilot programme designed to address concerns expressed in a national inquiry into the educational experiences and outcomes of children and young people with SEND (Lamb, 2009). The original intervention had three strands: (1) assessment, tracking, and intervention, (2) structured conversations with parents, and (3) provision for developing wider outcomes (such as behaviour and attendance). The pilot was conceptualised as one that required strong leadership and involved National Strategies' regional co-ordinators and advisory teachers working with link teachers to develop school action plans around the three strands for intervention. Schools were also provided with additional support from their local authorities (LAs) and link staff attended conferences organised by the National College for School Leadership (Department for Children, Schools and Families, 2009b, 2009a).

A two-year national evaluation of the AfA pilot involving around 12,000 pupils and over 400 schools across ten LAs found statistically significant improvements in teacher assessments of academic progress among students with SEND in both English and mathematics (Humphrey and Squires, 2010, 2011a, 2011b). Compared to national rates of progress among students with SEND, the effect sizes (the number of standard deviations of difference between the means) associated with these changes ranged from 0.17 to 1.39, but in all cases they were of sufficient magnitude to be considered practically meaningful (Barlow et al., 2015). There was also clear evidence of subgroup moderator effects—for example, academic progress varied significantly by stage of SEND provision (for example, School Action Plus) and primary need (for example, autism; Humphrey and Squires, 2011a). Parental engagement with schools also appeared to improve over the course of the pilot, although this finding was not statistically significant (Lendrum, Barlow and Humphrey, 2015). Finally, in terms of wider outcomes, attendance of persistent absentees improved by around 10% across the pilot sample (Humphrey and Squires, 2011a), and there were significant teacher-reported improvements in positive relationships, and reductions in bullying and behaviour problems when compared to a control group of pupils with SEND from schools not implementing AfA (Humphrey, et al., 2013).

Qualitative process data gathered from a subsample of 19 case study schools indicated that schools appreciated the flexibility inherent in AfA. It was most successful when schools built on existing practices and when the lead teacher was a member of the senior management team. Leadership was also found to be important in the implementation of structured conversations with parents. The project provided training and networking opportunities to develop provision for SEND pupils both within and across schools. Schools made better use of data than they had previously to inform target setting and monitoring of pupil progress in order to inform provision, teaching, and promote positive outcomes. The development of structured conversations was particularly helpful by providing a holistic view of pupils and by increasing parental engagement. Schools were determined to engage 'hard-to-reach' parents and made use of additional funding available to extend or adapt existing arrangements (Humphrey and Squires, 2011a).

Following the success of the AfA pilot, a national charity was formed in 2011—AfA 3As (Aspiration, Access, and Achievement). AfA 3As began to scale up the programme and is currently working with over 1,500 schools;[2] cumulatively, the programme has reached over 2,500 schools (Price Waterhouse Cooper, 2016). In the period since the formation of AfA 3As, the programme has undergone significant development, however it still retains some features of the pilot programme. The three original strands have been renamed but are essentially covering the same ground, while 'leadership and governance' now features as a distinct core element of the programme. The anticipated outcomes remain focused on literacy, numeracy, and wider outcomes (for example, parental engagement, attendance, engagement in learning, behaviour, and participation in extra-curricular school activity; Price Waterhouse Cooper, 2016). Schools remain able to develop action plans and select activities with guidance from an external AfA coach (Achievement for All 3As, 2015).

However, the current version of AfA also diverges from the national pilot in a number of important ways:

- AfA is now nationally rolled out rather than being arranged in clusters of schools. This means that the use of existing networks (for example, at the LA level) is not emphasised as part of the intervention and schools buy into the project individually.

- Most funding is from the school budget (either direct schools grant or pupil premium); other sources include donations from businesses, foundations, philanthropists, and LAs. A small number of schools cluster-fund (several schools each contribute a share of the cost of AfA). Eighty percent of schools report that the programme offers value for money (Price Waterhouse Cooper, 2016).

- As noted above, AfA now targets the 'lowest achieving 20%' of pupils as defined by each school and this is not limited to pupils with SEND. Price Waterhouse Cooper (2016) reported that the target population that AfA school champions wanted was mainly children in three categories: those described as 'not achieving', those in receipt of free school meals (FSM), and/or those in receipt of SEND support. AfA has also been extended to include all year groups, having only been focused on those in Years 1, 5, 7, and 10 in the pilot. In this sense, the target group in the current AfA intervention has diversified considerably from the national pilot.

- Having been exclusively delivered face-to-face in the national pilot, some key training is managed via an online space referred to as 'The Bubble'. At the time of writing it comprises more than 100 study modules containing over 100 hours of professional development opportunities. Staff and governing bodies in AfA schools receive individualised logins and there is support for three levels of professional engagement (from short awareness-raising sessions to Masters-level study).

- There is now a bespoke theory of change for the AfA programme; this was agreed as the focus of this evaluation at the start of the project and is available in Appendix A.

An RCT of the newer version of AfA was conducted as part of the National College for Teaching and Leadership's 'Test and Learn' programme in which the Centre for British Teachers, Centre for Use of Research and Evidence in Education, the Oxford Department of Education, and Durham University worked in partnership to support trial and programme set up (Churches, 2016). This two-year trial, involving approximately 2,200 pupils, found small but statistically significant

[2] Estimate provided by AfA 3As.

negative impacts on attainment at the intention-to-treat (ITT) level (*d* = -0.17) and at the FSM subgroup level (*d* = -0.43). Thus, to date there have been two independent evaluations of AfA with seemingly contradictory results.

Substantial methodological limitations of both studies may render their findings insecure. The pilot evaluation was not an RCT and relied on teacher assessment of academic outcomes instead of standardized tests. Furthermore, the pilot examined the AfA programme under ideal conditions, with schools benefiting from generous levels of funding, additional support from LAs, and oversight from the National Strategies and National College for School Leadership. Finally, as previously noted, the AfA programme has evolved significantly since the pilot. While the Churches (2016) RCT did reflect the newer version of the intervention, there were major issues in relation to attrition, the scale of which alone would render the trial as having *zero* padlocks in the EEF trial security classification system. Furthermore, although described as a two-year trial, it actually involved only a single year of direct comparison between AfA and business as usual as the control schools began to implement AfA in the second year. Finally, this trial lacked an implementation and process evaluation (IPE), and failed to conduct subgroup moderator analyses for the AfA target group. Thus, there is genuine equipoise regarding the impact of AfA, setting the stage for a definitive effectiveness trial.

Intervention

The AfA programme evaluated in this trial is a national school-based intervention 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 core areas: 'leadership for inclusion', 'teaching and learning', 'wider outcomes and opportunities', and 'engaging with parents and carers'. The direct recipients of AfA are the teachers in a school and it takes place in the participating school over a minimum two-year period. The following narrative description is derived from the Template for Intervention Description and Replication (TIDieR; Hoffmann et al., 2014).

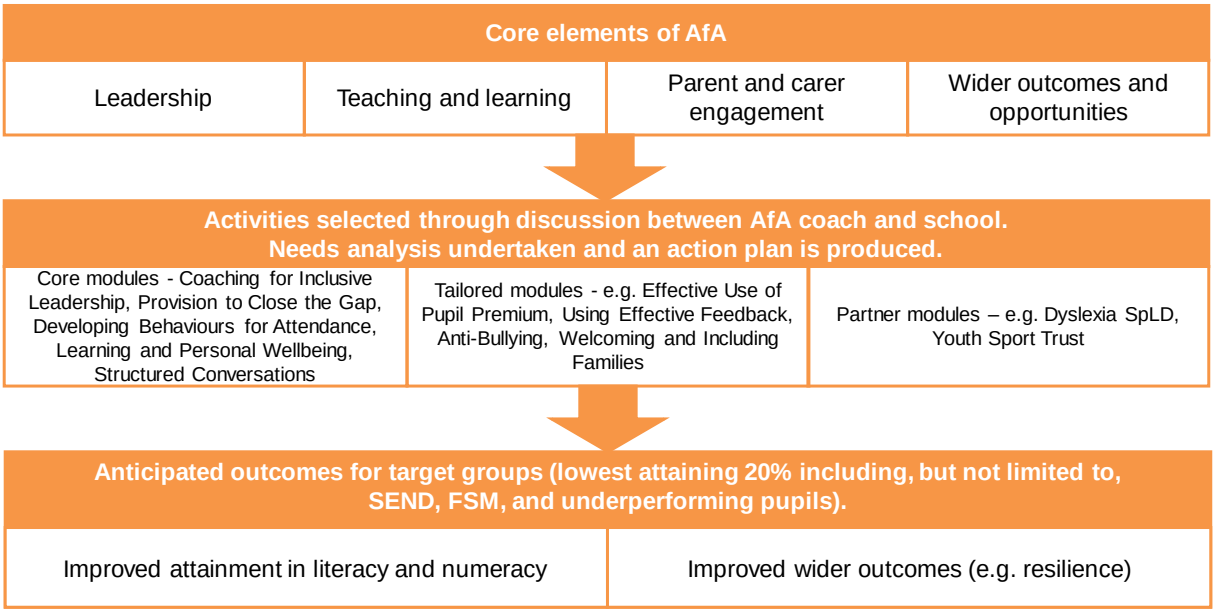
Each participating school had a designated member of staff known as the AfA champion working in collaboration with an AfA coach (employed by AfA 3As). They meet to discuss the needs of the school (using existing data held by the school as a basis for said discussion) and generate an action plan. This involves co-ordinating and selecting priority school/teacher development areas based on the four areas mentioned above. The four modules offered relating to each area each contains a core module and a list of 'tailored' modules. For example, the core area 'engaging with parents and carers' has one core module named 'Structured Conversations with Parents' (a process that provides parents with the opportunity to contribute to their child's learning and express their views through termly discussions with the class teacher; these discussions follow a structure of four key stages: explore, focus, plan, and review; Lendrum, Barlow and Humphrey, 2015), and two tailored modules called 'Early Support' and 'Welcoming and Including Families'.

AfA is designed to be flexible and is expected to be tailored to the specific needs and priorities of each participating school, which are agreed in the initial needs analysis conducted by the AfA coach with the school's AfA lead. Similarly, the '20% target group' is determined by the school and this means that the selection criteria may be interpreted differently in different schools (for example, those pupils for whom the most recent test data places them in the bottom 20%, or those that teachers consider vulnerable to underachievement). Thus, this group could be the lowest 20% of pupils based on assessment data or it could be another group that the school believes are vulnerable for other reasons (for example, those eligible for FSM, travellers, migrant children, those with SEND, or children in military families who tend to be highly mobile).

Each AfA school also had a lead teacher who usually acts as the AfA School Champion. The School Champion's role is to co-ordinate with all teachers within the school and is ideally a leadership role. In schools where there is an additional AfA lead, their role appears to be to introduce AfA to the school and then to link in with the senior management team (in cases where they are not a member of said team). A step-by-step guide is provided for participating schools by the AfA coach. Teachers are also able to access an online learning platform (The Bubble) containing the core, tailored (those selected by the AfA coach based on the school's priority development area), and partner modules (for example, 'BBC Children in Need Fun and Friendship'). Teachers can access these modules online via The Bubble to negotiate their own learning or can have the AfA coaches deliver them. Each AfA school gets up to 12 'interactions' with its AfA coach in a school year, supporting the school in its attempts to enact changes in practice in the specified areas. Coaching visits include bespoke activities that are pertinent at a given point in time (for example, an introduction to the programme with all staff in the first term of implementation), training in specific aspects of the intervention (for example,

structured conversation training), and termly review meetings. Schools taking part in the intervention also have the opportunity to work towards an AfA 'Quality Mark' (QM) based on the AfA coach evaluation of the progress made against a set of AfA-derived standards.[3] Not every school taking part in the AfA programme is awarded QM status. A basic schematic of the current AfA intervention is displayed in Figure 1 and the theory of change schematic agreed for evaluation at the start of the project is in Appendix A.

Figure 1: A simplified version of the current AfA programme



This theory of change contains several adaptations and compromises needed in order to evaluate AfA rigorously in a cluster RCT with accompanying IPE. First, in order to comply with the recommendation to specify subgroup analyses in advance, participating schools were required to nominate pupils for the target group prior to randomization using bespoke guidance provided by AfA 3As. Under normal circumstances, this would happen as part of the needs analysis that takes place once a given school has signed up to participate in the intervention (and indeed, the composition of said target group may subsequently change as schools review the progress being made by pupils). In this report, we have used the target group identified prior to the baseline of the trial in our analyses. Second, the inherent flexibility of AfA meant that the quantitative data-generation undertaken to examine progress in implementation as part of the IPE focused primarily on fidelity to key intervention *principles* as opposed to pre-specified *practices*. Third, we were mindful of the duration of intervention exposure and the impact that this might have on the achievement of intended outcomes given that AfA is designed to be implemented for two full school years (six terms). Hence, while this main report examines the impact of the intervention on academic outcomes of pupils exposed for around five terms, a forthcoming addendum report will examine its impact on pupils exposed for six or more terms.

Evaluation objectives

Our team conducted a large trial to evaluate the effectiveness of the AfA programme in England. We focused on the intervention's impact on pupils' academic attainment (for example, reading and mathematics) and wider, resilience-related outcomes. We also examined its impact on (1) the AfA target group and (2) those eligible for FSM. Furthermore, we investigated variability in the implementation of AfA and whether this influenced outcomes. The specific research questions are outlined below. The protocol and statistical analysis plan for the project can be accessed here: **https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/achievement-for-all/**.

[3] The AfA schools programme primary school quality scheme has two awards including (1) the Quality Mark (QM) and (2) the Quality Lead (QL). All schools that successfully completed the AfA programme were expected to achieve QM status. This was evaluated in the second year of the programme against AfA extended criteria. At the start of the trial, AfA estimated that around 50% of schools would achieve QM status over the course of this trial.

Research questions

1. Compared to business as usual, what is the impact of AfA on children's literacy (Year 5 whole group, primary outcome), mathematics, attendance,[4] and resilience-related outcomes (secondary outcomes)—
    a. after five terms of exposure—Year 5 (2016/2017) cohort?
    b. after six or more terms of exposure—Year 4 (2016/2017) cohort?
    c. What are the perceived impacts of AfA among intervention stakeholders (for example, teachers, headteachers)?

2. Are there differential intervention benefits in the above outcomes among pre-specified subgroups of children—
    a. those eligible for FSM?[5]
    b. the target group of children identified by participating schools as belonging to 'the lowest achieving 20%' (note that in the Y5 AfA target group, literacy is treated as co-primary outcome)?
    c. What processes underpin any differential intervention benefits identified?

3. 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?
    c. To what extent does contextual variation influence the implementation of AfA (and, subsequently, outcomes)?
        i. How and why is this the case?

4. Is there evidence to support the AfA theory of change?

Ethics and trial registration

This research was approved by the University of Manchester Research Ethics Committee (UREC) for both the RCT (reference: ethics/16292; approved 1 August 2016) and the IPE (reference: ethics/16414; approved 28 September 2016).[6] The RCT was registered with the ISCRTN (registry trial reference number: ISRCTN67347514; see **http://www.isrctn.com/ISRCTN67347514**). All schools involved in the project were recruited by AfA 3As, the organisation responsible for delivering the intervention, before the start of the trial.

Participating schools signed a Memorandum of Agreement (MoA) indicating their willingness to take part. A copy of this document is provided in Appendix C. The MoA provided schools with information regarding what participation in the AfA trial would entail (for example, the randomisation and allocation procedure and the data collection requirements of intervention and control group schools) and what they could expect in return for their participation.

Data protection

During the trial, data was collected on pupils' attainment and absence (provided by the National Pupil Database—NPD), socio-demographic information (for example, FSM eligibility, provided by the schools), and resilience-related outcomes (for example, school connection, acquired via online surveys). Pupils completed short surveys at two time periods: September/October 2016 and April/May 2018. These were conducted in schools through a secure online platform and responses were accessed by the University of Manchester. For the purpose of research, these responses were linked

[4] Among children in the AfA target group only ('the lowest achieving 20%').
[5] Specifically, EverFSM6 – e.g. eligible for FSM at some point in the last 6 years
[6] There were two different processes relevant to the participants' involvement in the project. One concerns General Data Protection Regulation (GDPR; see the 'Data Protection' section) and the other concerns research ethics. From a GDPR perspective, the legal basis for processing the data is not consent but public interest. We did however seek consent to participate from the pupils and their parents to fulfil our research ethics requirements. The requirements of ethics consent are not the same as the GDPR, they cover a wider remit, including, for example, protecting research participants from harm.

with information about the pupil from the NPD and shared with the Department for Education, the Education Endowment Foundation (EEF), the Fischer Family Trust (FFT—EEF's data processor for the archive), and, in an anonymised form, with the Office for National Statistics (ONS). The data will also, potentially, be shared with other research teams and further matching to NPD data may take place during subsequent research. Data was treated with the strictest confidence using pseudo-anonymised information in line with General Data Protection Regulation (GDPR).

At the start of the project (September/October 2016) parents/carers were provided with a privacy notice (participant information sheet, PIS, see Appendix D) outlining what would happen to the data or information relating to their child and what their rights were in relation to this data. At a later date during the trial (May/June 2018) an amended privacy notice was provided to each parent/carer (see Appendix E). This amended document was issued to provide further information that complies with recent changes in GDPR and the Data Protection Act 2018 (DPA, 2018). These forms also encouraged parents/carers to visit the following ICO websites for further information regarding data protection:

https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/ and **https://ico.org.uk/for-the-public/is-my-information-being-handled-correctly/**.

The University of Manchester was responsible for collecting and processing the data, was the organisation in control of personal data throughout the project, and was responsible for the production of this report. Following data archiving upon trial completion, the EEF became the data controller for the FFT Education archive. Identifiable information will not be transferred outside the EU or included in any outputs related to the project, and appropriate measures were taken to ensure the data remained secure at all times.

Pseudo-anonymised information—where individuals are not readily identifiable—was held during the four-year period that the research project was active, after which it was edited to ensure individuals in the data set are completely unidentifiable.

Data retention

This anonymous information is retained and may be used for research for another five years. After 2025, the information and data will be securely destroyed by the University of Manchester. The EEF has its own data retention policy, which can be accessed here: here:
https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/evaluator-resources/data-protection/.

The legal basis for processing data

The type of data that was used in the AfA trial evaluation included pupils' background information (for example, gender, SEND status, and FSM eligibility, collected from schools), survey data on resilience-related outcomes (for example, self-esteem and aspirations), attainment data (KS1 and KS2 literacy and mathematics scores obtained via the NPD), and attendance data (obtained via the NPD).

Under GDPR, the legal basis for processing personal data for the research project is *public interest*. It is in the public interest to raise the achievement and aspirations of all children, in particular of those facing socio-economic disadvantage. This project is also in the public interest because it will help determine whether the AfA programme is successful in achieving this goal by evaluating whether it can improve the attainment and wider outcomes of the lowest achieving 20% of pupils. This research will inform future educational provision.

The relevant articles are:

- Personal data: GDPR Article 6(1)(e). Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.
- Special categories of personal data: GDPR Article 9(2)(j). Processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

Project team

The evaluation was led by Prof Neil Humphrey and Dr 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. Dr Patricio Troncoso undertook additional quantitative analysis duties. Dr Ola Demkowicz supported data generation in the IPE strand of the trial and led the analysis of the qualitative data. Lawrence Wo undertook a number of data management duties. All of the above named staff worked for the University of Manchester during the trial.

At the request of AfA, the delivery team are not named individually in this report.

Methods

Trial design

Trial type and number of arms		Two-arm school-RCT
Unit of randomisation		Schools
Stratification variable(s) (if applicable)		%FSM, %SEN, %RWM
Primary outcome	variable	Reading attainment (Y5 cohort and Y5 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)	Reading attainment (Y4 cohort)* Mathematics attainment (Y5 cohort) Mathematics attainment (Y4 cohort)* Attendance of AfA target pupils (Y5 cohort) Resilience related outcomes (Y5 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 (Y5 cohort and Y5 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 (Y5 cohort) Mathematics attainment (Y4 cohort) Attendance of AfA target pupils (Y5 cohort) Resilience related outcomes (Y5 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

*NB: All outcomes marked with * will be addressed in the aforementioned addendum report.*

A two-arm cluster RCT design, incorporating a mixed-methods IPE, was used. Pupils are grouped by the school they attend into naturally occurring clusters and so school was used as the unit of randomisation. AfA is a whole-school intervention, meaning other forms of randomisation (for example, class- or year-level) were not feasible.

Schools were randomly allocated to one of two trial arms: (1) deliver the AfA programme (intervention arm) or (2) business as usual (control arm). See the Randomisation section for more details. In order to minimise differential attrition by trial arm, the schools allocated to the control arm were offered a retention incentive of £1,000 (to be paid in instalments of £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 trial period (2016/2017 and 2017/2018).

Participant selection

Participation was restricted to those schools that were not already involved—or never had been involved—with the AfA programme. The school recruitment process was carried out by AfA 3As which in the first instance targeted schools in the North East of England before widening the catchment area to other regions. Ultimately, participating schools 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.[7]

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 business as usual control arm (see the Randomisation section for more details). The target cohort within the schools were pupils in Year 4 (Y4) and Year 5 (Y5) in the first year of the trial (2016/2017). After accounting for parental/carer opt-outs (*n* = 117; 0.91%),[8] this sample consisted of *n* = 6,586 and *n* = 6,338 pupils in the Y4 and Y5 cohorts, respectively. Within the cohorts, 1,350 Y4 and 1,374 Y5 pupils were nominated as the target cohort (22% and 21%, respectively). The target group was identified for all schools prior to randomisation. Participating schools nominated their 20% target group as part of the process of signing up for the evaluation using guidance provided by AfA. The target population were identified jointly by schools and coaches as part of the normal AfA procedure based on joint decision-making.

Outcome measures

The analyses in this report focus on the outcomes of pupils who were in the Y5 cohort at the start of the intervention (2016/2017). The outcomes of pupils in the Y4 cohort will be assessed and reported as an addendum to the main trial report. As well as examining outcomes on the cohorts as a whole, we were also specifically interested in the impact of AfA on subgroups of pupils, specifically those in the AfA target group and those eligible for FSM.

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 2018) for pupils in the Y5 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 2018).

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). See Appendix F for a copy of the questions that the pupils were asked. 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

[7] 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.
[8] 63 and 54 pupils, in the Y4 and Y5 cohorts, respectively.

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 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 (2017/2018) was obtained for the members of this subgroup in the Y5 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%.[9] 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, 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, we used count data as our outcome measure for the attendance variable (that is, number of unauthorised absences for the 2015/2016 academic year and for the 2017/2018 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, meaning it is prone to bias. This decision is consistent with current EEF policy in relation to use of KS2 writing data.

Sample size

Calculation of sample size

Sample size calculations were based on our primary outcome measure and carried out using the Optimal Design programme (version 3.01). Initial calculations determined that 140 schools would be needed, yielding an estimated sample size of 4,800 pupils per cohort (an average cluster size of approximately 40 pupils per school after allowing for some attrition at the baseline data collection stage). Based on an intra-cluster correlation coefficient (ICC) of no more than 0.14,[10] a pre- to post-test correlation coefficient of at least 0.70 (R^2 = 0.49; based on EEF guidance), and standard power and alpha thresholds of 0.80 and 0.05, respectively, we calculated that the trial would be adequately powered for a minimum detectable effect size (MDES) of 0.15 in the ITT analysis (RQ1a; whole Y5 sample; average cluster size of approximately 40). This sample size would also be sufficiently powered for an MDES of 0.18 in the FSM sub-sample (RQ2a; estimated to be 30% of the sample with an average cluster size of approximately 12), and an MDES of 0.20 in the AfA target subgroup (RQ2b; estimated to be 20% of the sample with an average cluster size of approximately 8).

Ultimately 134 schools were recruited to the trial at the randomisation stage that met the eligibility criteria as per the MoA. Despite not achieving the target of recruiting 140 schools, the trial was more than adequately powered (see MDES estimates of the achieved sample displayed in Table 2 below). All MDES calculations were calculated for the Y5 (2016/2017) cohort only, upon which this report is based.

[9] 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.
[10] Calculated using KS1 Literacy Point scores obtained from NPD data.

A total of 12,924 pupils participated in the trial (Y5: *n* = 6,338; Y4: *n* = 6,586). Of this sample, 1,374 Y5 and 1,350 Y4 pupils were nominated as the target AfA cohort (22% and 21%, respectively). As already noted, this report is based on the Y5 (2016/2017) dataset, with the analysis of the Y4 (2016/2017) cohort to be included as an addendum report.

Table 2: MDES estimates for the achieved Y5 (2016/2017) sample

Cohort	No. of clusters	Average cluster size	Pre- to post-test correlation (R^2)	ICC	Power	*p*	MDES
Whole	134	47	0.70 (0.49)	0.08	0.80	0.05	0.14
FSM	129	14	0.70 (0.49)	0.08	0.80	0.05	0.16
AfA target	134	10	0.70 (0.49)	0.08	0.80	0.05	0.17

Note: The MDES here is that at randomisation stage, calculated using the 'mdesapp' online calculator (Troncoso, 2020), which uses variance explained by the baseline covariate at levels 1 and 2 ($\boldsymbol{R_1^2}$ and $\boldsymbol{R_2^2}$). See Table 5 for further details.

Randomisation

Randomisation took place in November 2016 following completion of the baseline pre-test surveys for the secondary resilience-related outcome measures. The randomisation procedure was conducted independently of the evaluation team by the Manchester Clinical Trial Unit to eliminate selection bias. A minimisation algorithm was applied utilising the following school-level co-variates sourced from EDUBASE:[11] %FSM, %SEN, and %KS2 reading, writing, and mathematics combined level 4+(RWM+4).[12] Given the nature of the AfA intervention and the primary trial outcome, these were deemed important variables on which to obtain balance at baseline. As a result of this randomisation process, and in order to achieve balance based on the minimisation protocol, 66 and 68 schools were randomly allocated to the AfA intervention and business as usual control arms, respectively.

Statistical analysis

ITT complete case analyses were conducted for all primary and secondary outcome measures using raw complete case data to address RQ1a. 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, if the intervention was only partially implemented in a given school). Multilevel models (MLM) with fixed slopes and random intercepts were performed for all outcome measures, with the exception of attendance. To account for the nested nature of the data, multilevel (hierarchical) 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).

Next, to investigate possible differential intervention benefits among pre-specified subset of pupils (RQ2a and RQ2b), additional subgroup analyses were performed for each of the primary and secondary outcome measures. The MLM model (as described above) was re-run twice with a subset of the main data – first, only FSM pupils were included in the analysis (Model 2.1), and second, only AfA target pupils were entered (Model 2.2).

Following principal analyses using fully observed data (that is, complete case analysis; Model 1.1), as per 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 indicator[13] were entered as co-variates. At the pupil level, SEND and gender were added as co-variates.

[11] Information missing from EDUBASE was sourced directly from the school.
[12] Attainment (that is, the proportion of pupils attaining level 4+ in reading, writing, and mathematics).
[13] Derived from the Usual Practice Survey (UPS) to provide a more robust estimate of the achieved relative strength of AfA.

For the models described above, those related to the resilience-related outcome measures were estimated using Mplus (version 8.2) in accordance with the SAP, whereas those based on data sources from the NPD were estimated in MLwiN (version 2.22; 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), the full information maximum likelihood (FIML) or multiple imputation (MI) procedures were used to re-estimate models to deal with missing data (whereby partially 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-variates (Model 1.2), were repeated using FIML or MI. For the resilience outcome measures these were estimated using FIML in Mplus (version 8.2) in accordance with the procedure described in the SAP. However, those related to outcome measures based on data sourced from the NPD 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 any intervention effects were present. Hedges' *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.[14] 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

MLMs were estimated using random intercepts and fixed slopes as planned for all outcome measures except attendance. For this 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. The dependent variable of post-trial absence was modelled with 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. Models estimated using data sourced from the NPD (relating to attainment and attendance) were performed using MLwiN (version 2.22) and the *lme4* (version 1.1-21) package in R (version 3.6.1) instead of MPlus. This deviation is due to the ONS not being able to provide access to MPlus software on the Secure Research Service. For the remaining outcome measures (the resilience-related measures), models were estimated using the MPlus (version 8.2) package as planned. FIML was used to account for missing data for resilience-related outcomes as planned. However, for outcomes where data was sourced from the NPD, models were re-estimated using joint modelling MI, 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. The SEM (multilevel path) analyses were also estimated in R (version 3.6.1) using the package *lavaan* (version 0.6-4) instead of MPlus.

Additional deviations relate to the statistical tests planned to address RQ3(i). Models that would have included *on-treatment* status as a school-level explanatory variable have not been conducted. This variable was to be derived from data provided by AfA 3As identifying which schools were awarded Quality Mark (QM) status by the end of their participation in the programme.[15] Schools awarded a QM were to be coded as *on-treatment,* while schools that did not achieve the QM would be coded as *off-treatment*. In order to assess whether outcomes vary as a function of *on-treatment* status, we planned to conduct a two-level MLM with only the data of intervention schools. Pre-test outcome (KS1 reading and writing scores) and QM status were to be entered as explanatory variables at the pupil level and

[14] Models with and without the group predictor were estimated and compared using the chi-square difference (likelihood ratio) test to provide additional confirmation of whether the predictor variable provides significant explanatory power to the model. These tests were not outlined in the SAP and do not affect the planned statistical tests. They have been added to provide additional information on whether the change in deviance is significant between models when an explanatory variable (for example, group) is added.

[15] See note 3, page 9.

school level, respectively, with KS2 reading entered as the response variable. Next, to address RQ3(ii), subgroup membership (first FSM eligibility and then AfA target group membership) was to be fitted as a pupil-level explanatory variable in order to model cross-level interactions (*QM status*sub-group membership).* This was to determine whether differential intervention benefits among specified subgroups vary as a function of *on-treatment* implementation. Before the start of the trial (summer 2016) AfA 3As estimated that approximately 50% of the intervention schools (approximately 33 schools) would achieve a QM status (based on information of participating schools at the time), however only 12 intervention schools ultimately achieved this award. A minimum of 30 groups is recommended for level-two sample size in multilevel modelling (Kreft, 1996; Maas and Hox, 2004; Hox, 2010); for samples with less than 20 groups, the fixed parameter estimates and standard errors become inaccurate (Hox, 2010). For this reason, we have not conducted these analyses as planned (see the Quantitative Implementation Data Analysis subsection below).

Effect size calculations

Effect sizes were calculated using Hedges' *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:

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

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

Following acquisition of post-test scores, we calculated the proportion of missing data for the intervention and control arms for each 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 FIML (in MPlus, version 8.2) or MI (in R, version 3.6.1).

We also examined missing data by conducting a regression analysis with a binary variable for complete and incomplete cases (coded as ones and zeros, respectively). A complete case was defined by the presence of both the pre-test (KS1 reading and writing) and post-test (KS2 reading) scores. The binary variable was entered into the regression analysis as the outcome variable, with condition (treatment or control), FSM (if eligible), and AfA target cohort (if AfA target) entered as explanatory variables. This allowed us to determine the likelihood of students having complete cases—and whether this is influenced by trial group membership (treatment versus control) or subgroup membership (whether eligible for FSM, or whether in the AfA target group).

Implementation and process evaluation

The IPE strand of the project comprised four distinct phases. In phase one, we surveyed participating schools about their usual practice at baseline (pre-randomisation) and follow-up (between April and July 2018). In phase two, we performed a quantitative analysis to identify and select eight schools using the baseline data obtained from schools in the RCT strand. Phase three involved a longitudinal case study analysis of the eight schools for the duration of the trial. Interviews with teachers and AfA champions within the school and with AfA coaches were used as the unit of analysis to help us to understand the extent to which teachers applied the principles of the AfA programme in their own practice. In the final IPE phase, which took place towards the end of the second year of the trial (between April and July 2018), a school-level implementation survey was administered to each school allocated to the intervention arm. The survey was developed during the first year of the trial using information derived from qualitative data collected during phase three.

The various IPE methods and their role in the study are summarised in Table **3**: IPE methods.

Table 3: IPE methods

Research methods	Data collection methods	Participants / data sources	Data analysis methods	Research questions addressed	Implementation/ logic model relevance
Academic attainment assessment (KS1 literacy)	NPD extraction	Y4 and Y5 pupil cohorts (6,160) in AfA schools (66)	Multilevel modelling	RQ3, RQ4	Context
Qualitative case studies	Interviews	AfA coaches (8), school AfA champions (9; 7 were deputy headteachers, 1 was a headteacher), headteachers (6; 1 was an AfA Champion), SENCOs (3), class teachers (7), AfA Link teachers not designated as AfA Champions (2) 26 people in total (2 had 2 roles) through 95 interviews	Thematic analysis	RQ3, RQ4	Context, implementation, and factors affecting implementation
Implementation surveys	Online surveys	School AfA leads (66, of which 53 completed)	Descriptive statistics, multi-level modelling (including structural equation modelling)	RQ3, RQ4	Implementation
Usual practice surveys	Online surveys	Nominated school contact (134, of which 134 completed at DP1)	Descriptive statistics, exploratory factor analysis, multilevel modelling	RQ1, RQ2, RQ3, RQ4	Context, usual practice, control group activity

Usual practice (and control group activity)

Usual practice surveys were administered to nominated contacts in participating schools prior to randomisation (September to November 2016; data point one, DP1) and during the second year of the trial (April to June 2018; data point two, DP2). At DP1, schools were asked to report what continuing professional development (CPD) had taken place since September 2014, and at DP2 they were asked to update this to also cover the trial period. At DP1, all 134 schools in the sample completed the survey, and at DP2 this dropped to 110 (55 schools in both groups). A copy of the DP1 Usual Practice Survey (UPS) is shown in Appendix K.

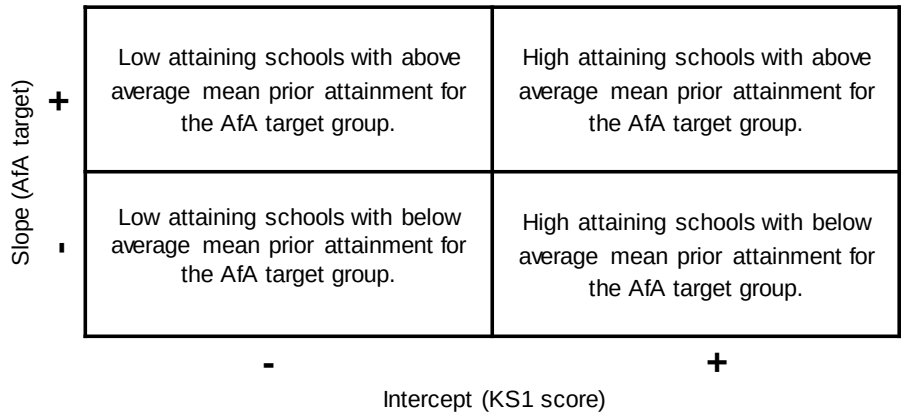
Case study selection process

The case study selection process took place following randomisation and prior to the first school visits. Hierarchical modelling of the baseline data of schools obtained as part of the RCT strand was performed to identify case study schools for phase one of IPE strand. We applied a two-level (pupils clustered in schools) random intercept and slope model using data from the AfA intervention schools only. KS1 literacy (reading and writing combined) was entered as the outcome variable along with a number of control variables (FSM eligibility, SEND, AfA target group) to calculate residuals for the schools.

Using the residuals, the intervention schools (n = 66) were then ranked as ‘low attaining’ or ‘high attaining’ depending on whether the school had a greater proportion of pupils with below- or above-average KS1 literacy scores, respectively. They were also ranked as ‘adding more value’ or ‘adding less value’ to the AfA target group depending on whether the

school had a greater proportion of AfA target pupils with above- or below-average KS1 literacy scores, respectively. By ranking on these two dimensions, we were able to allocate each school to one of four quadrants in a covariance matrix plot (see Figure 2). Schools were ranked within each quadrant and we then chose an equal number from each to be included in the case study phase in order to provide variability across our case study schools in terms of attainment levels and value added for the AfA target group.

Figure 2: An illustrative example of the residual covariance matrix plot



Qualitative case studies

We conducted longitudinal case studies of a sample of eight AfA case study schools. Five termly visits were made to schools to interview participants using semi-structured interviews (see Appendix B).[16] Some participants had multiple roles, especially in smaller schools, and the interviews were targeted at AfA coaches (39 interviews), school AfA leads (38 interviews), headteachers who were not AfA leads (3 interviews), SENCos who were not AfA leads (5 interviews), and class teachers (10 interviews). This data was used to answer a range of process-related research objectives. Our primary interest was how well schools were able to implement AfA and the analysis was driven by concepts based on implementation research (see below; Durlak and DuPre, 2008; Lendrum and Humphrey, 2012). Data was transcribed and entered into a qualitative analysis software package, NVivo (version 12).

To understand how the AfA intervention was implemented in the case-study schools, we focused on a number of dimensions, including:

- fidelity and application (How well did schools adhere to the principles of AfA while at the same time adapting AfA to their own circumstances?);
- dosage (In what quantity were aspects of AfA implemented, and were activities related to AfA implemented as often as they should have been according to the programme specification?);
- reach and responsiveness (Given that AfA is a school improvement programme that also has a specific group of pupils in mind, how well does AfA fulfil these dual roles?);
- quality (What approaches were used to strengthen the use of the AfA programme?); and
- programme differentiation (What distinguishes the AfA programme from other approaches that schools use to support a target group of pupils?).

We were also interested in understanding factors affecting implementation, including

- programme systems and characteristics;
- the influence of school factors;
- the influence of teacher factors;
- perceptions of the impact of AfA;

[16] Some information was also gathered via field notes during informal observations and from school documentation in order to provide additional context in support of our analyses; however, only interview data was included in the qualitative analysis.

- parental engagement (given that increasing engagement with and of parents through structured conversations is a central element in AfA, we examined what affects how well this was implemented); and
- sustainability (To what extent will schools continue with AfA or the approaches used once the trial has been completed?).

Implementation survey

An implementation survey was designed using knowledge gained from an Intervention Delivery and Evaluation Analysis (IDEA) workshop undertaken at the outset of the trial, and from the first three terms of the longitudinal implementation case studies, to investigate how intervention schools implemented AfA and whether any variability in implementation predicted intervention outcomes (RQ3). A copy of this survey is provided in Appendix G: The AfA Teacher Implementation Survey. The survey was administered through a secure online portal during the second year of the trial (between April and July 2018) and completed by the school's AfA lead. Data generated from this survey was used to construct the implementation variable(s) (see below).

Quantitative implementation data analysis

To address RQ3b, only data from schools involved in the intervention arm of the trial was used (meaning 'group' was not entered as a predictor variable into the following models). Two-level MLMs (pupils clustered in schools) with fixed slopes and random intercepts were fitted for the primary outcome measure (KS2 reading). Pre-test (KS1) scores were entered as a pupil-level explanatory variable, while the implementation variables were added as school-level predictors (resulting in Model 3.1). This was to determine the association between implementation and outcome variability.

Due to the lack of the QM status variable (see the Deviations from the SAP subsection above), models for the quantitative implementation data analysis were collapsed into the following: first, Model 3.1 was estimated as planned; next, FSM was entered into Model 3.1 to model cross-level interactions between the implementation variables and FSM eligibility, resulting in Model 3.2; then, AfA target group membership was entered into Model 3.1 to model cross-level interactions between the implementation variable and AfA target group membership resulting in Model 3.3. Changes in the log-likelihood and VPC are reported to show whether the introduction of these variables or the interaction explains any further variation in the model.

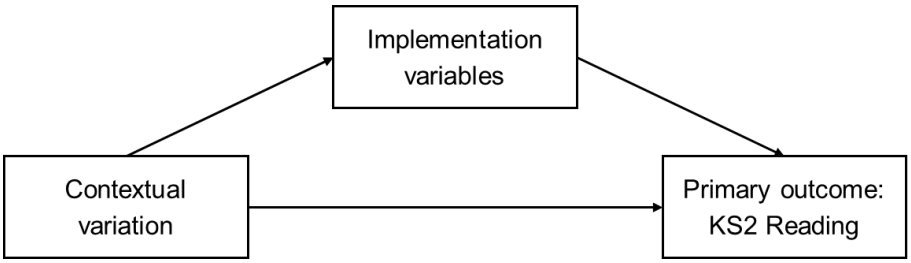
To address RQ3a (specifically, why AfA implementation varies), RQ3c, and RQ4, we performed two-level (pupils clustered in schools) structural equation modelling (SEM), specifically multilevel path analysis in R (Version 3.6.1) using the *lavaan* package (Version 0.6-4). SEM was used to examine the relationship between contextual variables and the primary outcome and also to determine whether this relationship is mediated by implementation variables (see Figure 3). The contextual variation variables used were the school-level randomisation parameters (%FSM, %SEN, %RWM 4+ categorized as 'low', 'medium', and 'high'), the school-level implementation variables were derived from data generated using the implementation survey (see below), and the pupil-level outcome measure was KS2 reading scores. Pre-test attainment scores (KS1 reading and writing combined scores) were also regressed on the outcome at both levels to account for baseline performance.

Data generated from the implementation survey was used to derive the implementation variables (see Appendix G: The AfA Teacher Implementation Survey for a copy of this survey). Fifty-three of the sixty-six intervention schools completed the survey. The flexible nature of the AfA intervention meant that different schools engaged with different aspects of the programme, and schools only answered survey questions relating to the particular modules that they engaged with (that is, those that were selected during the needs analysis phase at the start of the trial).[17] This means there are variable rates of response for the different sections of the survey. Due to the lack of complete data across the survey questions, we were unable to use exploratory factor analysis to reduce the data into underlying latent factors. Therefore, we averaged the responses within each category of the questionnaire—categories that represented the core modules of the AfA programme—resulting in variables for 'leadership and inclusion', 'teaching and learning', and 'provision and wider outcomes'. Given the proposed importance of the structured conversations aspect of the programme (Humphrey and Squires, 2011a, 2011b), two variables were utilised for this component: 'structured conversations fidelity' (average response rating for questions pertaining to the fidelity of the structured conversations component) and 'structured

[17] With the exception of the structured conversations aspect of the AfA programme—a component which all intervention schools were required to engage with as part of the intervention. Therefore, all 53 intervention schools that responded to the survey answered the questions for this component.

conversations dosage' (relating to the amount of structured conversations a school conducted per target group pupil throughout the academic year).

Figure 3: Model specification (relationships that were explored)



The findings from this analysis allowed us to investigate the extent to which contextual variation influenced the implementation of AfA (RQ3c) and to establish whether there is evidence to support the AfA theory of change (RQ4) since this is underpinned by an assumption that higher levels of implementation drive improvements in pupil outcomes.

Qualitative implementation data analysis

Qualitative data was analysed using thematic analysis in line with Braun and Clarke's (2006) six-step process: familiarisation with the data, generation of initial codes, searching for themes across codes, reviewing and refining of themes, defining and naming of themes, and generation of the report. This was undertaken as a hybrid analysis whereby the analytical process is guided by existing understandings (deductive) but also allows for the emergence of unanticipated themes relating to the research question (inductive). Specifically, coding and theme development was guided by existing evidence and literature around programme implementation, particularly that of Durlak and DuPre (2008) and Fixsen et al.'s (2005) reviews of implementation research as well as guidance from Humphrey et al. (2016). In line with such research, analysis was undertaken with attention to implementation dimensions (for example, fidelity and adaptation), factors affecting implementation, perceptions of impact, and sustainability. Within this overarching framework, emergent themes were developed based on the specific points raised by participants (for example, the particular factors influencing implementation).

Costs

Basic delivery costs were obtained from AfA. We also collected more detailed cost information from participating schools involved in the case study implementation evaluation via interviews with staff. This enabled us to estimate additional resources and teacher time required to implement the AfA programme over the trial period.

Timeline

The timeline for the project is outlined below in Table 4.

Table 4: AfA trial project timeline

Date	Activity
April–July 2016	Trial school recruitment
June 2016	IDEA workshop
July 2016	School and pupil background data acquired
August 2016	Ethical approval granted for RCT strand
September 2016	Ethical approval granted for IPE strand
September–October 2016	Pre-test Student Resilience Survey completed
September–November 2016	Usual practice survey—time one
October 2016	NPD request for KS1 data (for Y4 and Y5 cohorts)
November 2016	Randomisation
December 2016	NPD KS1 data provided (for Y4 and Y5 cohorts)
December 2016–January 2017	Case study school selection
January 2017	Initial training for intervention schools
January 2017	Implementation of AfA programme begins
March–December 2017 and January–July 2018	Data collection in case study schools
April–June 2018	Usual practice survey—time two
April–July 2018	Post-test Student Resilience Survey completed (for Y4 and Y5 cohorts)
April–July 2018	Implementation survey completed
July 2018	Y5 cohort leaves participating schools
December 2018	Implementation ends
October 2018–March 2019	Qualitative IPE analysis and write-up
January 2019	NPD request for KS2 and absence data submitted (for Y5 cohort)
June 2019	Post-test outcome data (NPD KS2 attainment and attendance variables) provided in the Secure Research Service (for Y5 cohort)
June–October 2019	Data analysis and report writing

Impact evaluation

Participants

One hundred and thirty-four schools (6,338 pupils) were recruited for the trial.[18] Of this sample, 66 schools (3,027 pupils) were allocated to the intervention arm and 68 (3,311 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 5,897 pupils (93%). At post-trial, outcome data (KS2 reading mark) was available for 6,029 pupils (95%). 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 (where both the pre-trial KS1 and post-trial KS2 scores were present) was available for 5,813 pupils (92%). The sample size of our complete case analysis (Model 1.1) yielded power for an MDES of 0.12 (see Table 5 for a summary of the MDES values for the different stages of the trial). See Figure 4: Participant flow diagram for Y5 (2016/2017) pupils for the participant flow diagram.

Table 5: Minimum detectable effect size at different stages of the trial

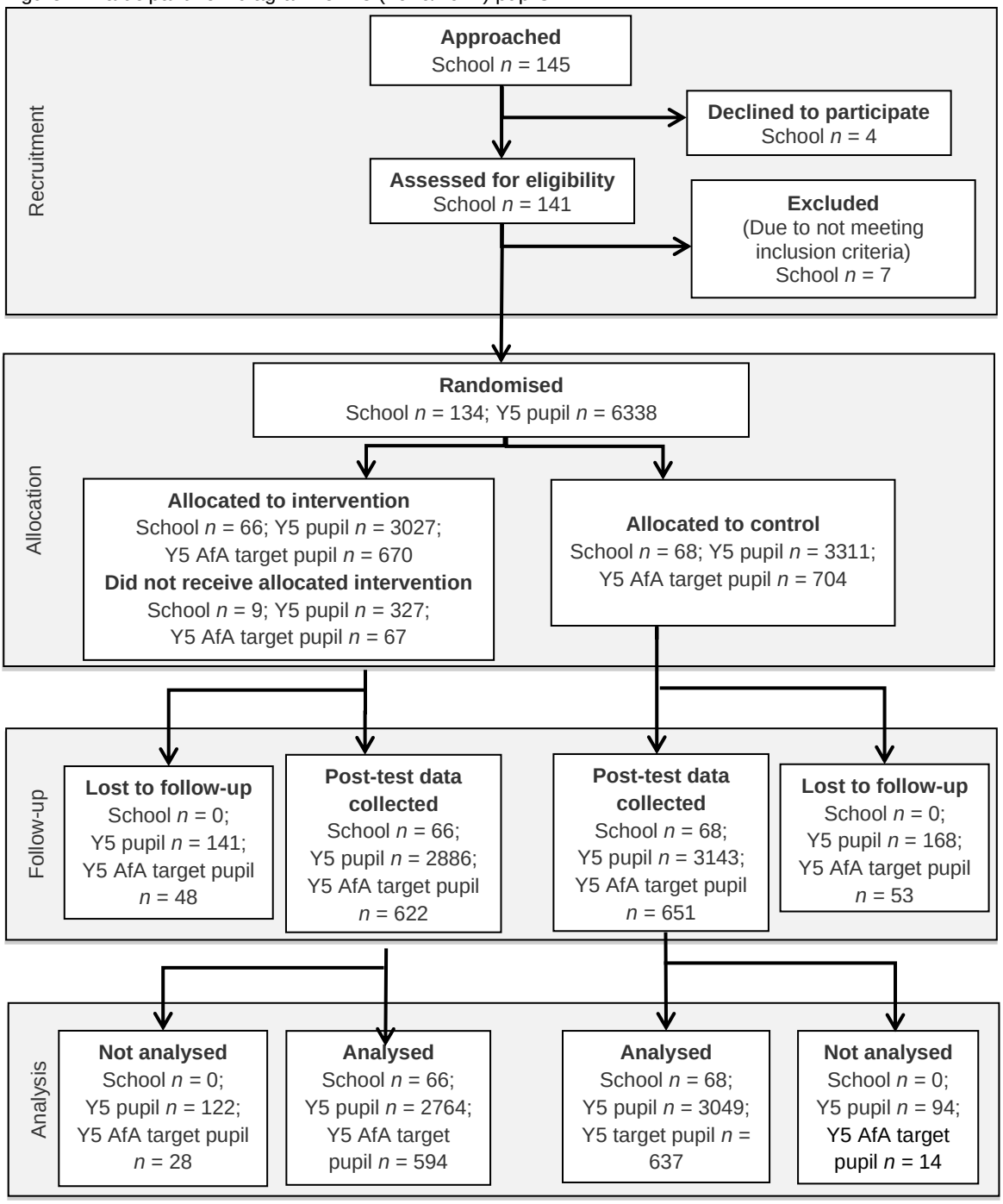
		Protocol			Randomisation			Analysis		
		Overall	FSM	AfA target	Overall	FSM	AfA target	Overall	FSM	AfA target
MDES		0.15	0.18	0.20	0.14	0.16	0.17	0.12	0.15	0.19
Pre-post test correlation (R)		0.49	0.49	0.49	--	--	--	--	--	--
R-squared		0.7	0.7	0.7	--	--	--	--	--	--
R_1^2	Level 1 (pupil)	--	--	--	0.47	0.41	0.41	0.465	0.454	0.29
R_2^2	Level 2 (school)	--	--	--	0.16	0.03	0.03	0.382	0.065	0.448
ICC	Level 2 (school)	0.14	0.14	0.14	0.08	0.08	0.08	0.086	0.062	0.171
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	47.3	14.4	10.3	43.4	13.4	9.2
Proportion randomised to treatment		0.5	0.5	0.5	0.49	0.50	0.49	0.49	0.49	0.49
Number of schools	Intervention	70	70	70	66	64	66	66	63	66
	Control	70	70	70	68	65	68	68	65	68
	Total	140	140	140	134	129	134	134	128	134
Number of pupils	Intervention	2800	840	560	3027	882	670	2764	820	594
	Control	2800	840	560	3311	973	704	3049	893	637
	Total	5600	1680	1120	6338	1855	1374	5813	1713	1231

Note: MDES = minimum detectable effect size; ICC = intraclass correlation coefficient.

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),[19] 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.

[18] This represents the sample size of the Y5 cohort only.
[19] 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**

Figure 4: Participant flow diagram for Y5 (2016/2017) pupils



Attrition

Attrition was 0% at the school level and 8.28% at the pupil level (8.69% and 7.91% for the intervention and control arms of the trial, respectively) based on a ratio of 6,338 randomised to 5,813 analysed for the primary trial outcome.

Table 6: Pupil-level attrition from the trial (primary outcome: Y5 reading, whole group)

		Intervention	Control	Total
N pupils	Randomised	3027	3311	6338
	Analysed	2764	3049	5813
Pupil attrition (from randomisation to analysis)	Number	263	262	525
	Percentage	8.69%	7.91%	8.28%

Table 7: Pupil-level attrition from the trial (primary outcome: Y5 reading, AfA target group)

		Intervention	Control	Total
N pupils	Randomised	670	704	1374
	Analysed	594	637	1231
Pupil attrition (from randomisation to analysis)	Number	76	67	143
	Percentage	11.34%	9.52%	10.41%

Pupil and school characteristics

Table 8**:** Baseline comparison of school and pupil characteristics of the Y5 whole cohort summarises the baseline school- and pupil-level characteristics of the whole Y5 cohort (134 schools; 6,338 pupils), and Table 9: Baseline comparison of pupil characteristics of the Y5 AfA target subgroup shows the baseline pupil-level characteristics of the Y5 AfA target subgroup. These tables include demographic, attainment, and resilience-related information.

Table 8: Baseline comparison of school and pupil characteristics of the Y5 whole cohort

Variable	Intervention group (*N* = 66)		Control group (*N* = 68)		
School-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage	
School Type:					
Academy Converted	14 (0)	21.2%	16 (0)	23.5%	
Academy Sponsored	9 (0)	13.6%	7 (0)	10.3%	
Community School	34 (0)	51.5%	30 (0)	44.1%	
Foundation School	1 (0)	1.5%	2 (0)	2.9%	
Voluntary Added School	5 (0)	7.6%	8 (0)	11.8%	
Voluntary Controlled School	3 (0)	4.5%	5 (0)	7.4%	
Ofsted rating:					
Outstanding	5 (0)	7.6%	9 (0)	13.2%	
Good	47 (0)	71.2%	47 (0)	69.1%	
Requires Improvement	13 (0)	19.7%	9 (0)	13.2%	
Inadequate	1 (0)	1.5%	3 (0)	4.4%	
Location/Setting:					
Urban	53 (0)	80.3%	56 (0)	82.4%	
Rural	13 (0)	19.7%	12 (0)	17.6%	
	Intervention group (*N* = 66)		Control group (*N* = 68)		
School-level (continuous)	n (missing)	Percentage	n (missing)	Percentage	
Proportion eligible for FSM	66 (0)	18.2%	68 (0)	17.6%	
Proportion with SEND	66 (0)	13.3%	68 (0)	14.7%	
Proportion achieving level 4+ in reading, writing, and mathematics (RWM+4)	64 (2)	79.7%	68 (0)	79.6%	
	Intervention group		Control group		
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage	
Proportion eligible for FSM (EverFSM 6)	882/3027 (0)	29.1%	973/3311 (0)	29.4%	
Proportion with SEND (S, E, or K)	513/3027 (20)	16.9%	622/3311 (0)	18.8%	
Proportion of male pupils	1560/3027 (0)	51.5%	1700/3311 (0)	51.3%	
	Intervention group (*N* = 3027)		Control group (*N* = 3311)		
Pupil-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Effect size
KS1 reading and writing combined	2875 (152)	15.49 (3.59)	3195 (116)	15.68 (3.65)	-0.05
KS1 mathematics	2875 (152)	15.91 (3.46)	3195 (116)	16.15 (3.47)	-0.07
Self-esteem	2782 (245)	11.72 (1.90)	3035 (276)	11.71 (1.85)	0.01
Goals and aspirations	2769 (258)	8.31 (1.67)	3019 (292)	8.17 (1.75)	0.08
Family connection	2705 (322)	17.91 (2.54)	2947 (364)	17.83 (2.52)	0.03
School connection	2693 (334)	17.24 (2.98)	2966 (345)	17.23 (2.83)	0.01

Note. The values for the pupil-level baseline measures have been calculated using all available data for a particular variable (that is, they are not based on complete case data) as per EEF guidance.

Table 9: Baseline comparison of pupil characteristics of the Y5 AfA target subgroup

	Intervention group		Control group		
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage	
Proportion eligible for FSM (EverFSM 6)	426/670 (0)	63.6%	428/704 (0)	60.8%	
Proportion with SEND (S, E, or K)	320/670 (5)	47.8%	326/704 (0)	46.3%	
Proportion of male pupils	400/670 (0)	59.7%	402/704 (0)	57.1%	
	Intervention group (*N* = 670)		Control group (*N* = 704)		
Pupil-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Effect size
KS1 reading and writing combined	637 (33)	12.76 (3.45)	683 (21)	13.12 (3.44)	-0.10
KS1 mathematics	637 (33)	13.58 (3.45)	683 (21)	13.97 (3.41)	-0.11
Absence at pre-trial[20] (2015/2016)	649 (21)	4.62 (9.55)	650 (20)	5.46 (12.35)	-0.08
Self-esteem	609 (61)	11.32 (1.93)	634 (70)	11.19 (1.98)	0.07
Goals and aspirations	602 (68)	8.05 (1.74)	632 (72)	7.90 (1.91)	0.08
Family connection	593 (77)	17.50 (2.76)	611 (93)	17.40 (2.79)	0.04
School connection	588 (82)	16.98 (3.01)	621 (83)	17.12 (3.00)	-0.05

Note. The values for the pupil level baseline measures have been calculated using all available data for a particular variable (that is, they are not based on complete case data) as per EEF guidance.

Imbalance at baseline

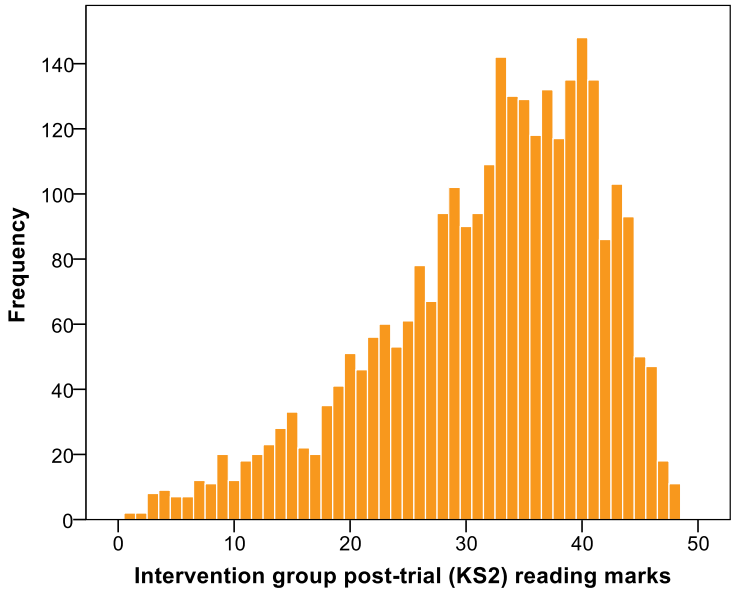
Effect sizes for the differences in scores between the intervention and control groups on pupil-level variables at baseline were very small for the whole Y5 cohort (KS1 reading and writing points = -0.05; KS1 math points = -0.07; self-esteem = 0.01; goals and aspirations = 0.08; family connection = 0.03; school connection = 0.01) demonstrating good balance and successful randomisation. Of particular note in relation to threats to internal validity and the security of trial findings is that the imbalance of our key observable at baseline (KS1 reading and writing combined) was -0.05, indicating low risk of confounding. Similarly, the effect sizes for the Y5 AfA target cohort subgroup were also small (KS1 reading and writing points = -0.10; KS1 math points = -0.11; pre-trial absence = -0.08; self-esteem = 0.07; goals and aspirations = 0.08; family connection = 0.04; school connection = -0.05).

Figure 5 illustrates the distribution of test scores for the primary outcome measure (KS2 reading). The distributions of the intervention and control groups are similar. The reading scores of both groups are negatively skewed (intervention = -0.776, SE = 0.046; control = -0.962, SE = 0.044), indicating the data has a long left tail and that most scores fall to the higher end of the distribution. The kurtosis of the reading scores of the intervention group is normal (0.132, SE = 0.091), but kurtosis is greater than normal for the control group (0.542, SE = 0.087), suggesting the reading scores have a relatively peaked distribution. Histograms of the baseline (KS1 reading and writing combined) scores are shown in Appendix H.

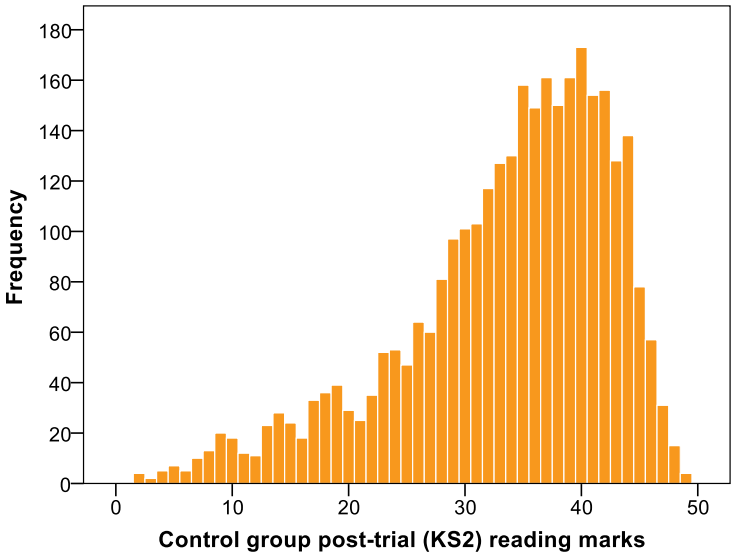
[20] This relates to children in the AfA target group only ('the lowest achieving 20%'). The original plan was to use a binary classification of persistent absentee versus non-persistent absentee for the attendance variable. Instead, we used count data of number of unauthorised absences for a given academic year (that is, over six terms during 2015/2016 for pre-test and over six terms during 2017/2018 for post-test).

Figure 5: 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. Cases for counts equal to one have been removed from the histograms to ensure they remain non-disclosive for individual pupils in the trial (in accordance with guidance provided by the DfE NPD and Data Sharing team and the ONS).

A.



B.



Outcomes and analysis

Basic descriptive statistics for the pupil-level outcome measures at pre- and post-test are displayed in Table 10 for the whole Y5 cohort and in Table 11 for the Y5 AfA target cohort subgroup.

Table 10: Means and SDs of pupil outcomes at pre-test (baseline) and post-test for the whole Y5 cohort

	Intervention		Control	
Outcome	Pre-test mean (SD)	Post-test mean (SD)	Pre-test mean (SD)	Post-test mean (SD)
KS1 reading and writing combined	15.49 (3.59)	-	15.68 (3.65)	-
KS2 reading	-	31.93 (9.47)	-	33.53 (9.15)
KS1 mathematics	15.91 (3.46)	-	16.16 (3.47)	-
KS2 mathematics	-	73.37 (24.56)	-	77.58 (23.37)
Self-esteem	11.72 (1.90)	11.95 (1.81)	11.71 (1.85)	11.91 (1.76)
Goals and aspirations	8.31 (1.67)	8.41 (1.67)	8.17 (1.75)	8.28 (1.68)
Family connection	17.91 (2.54)	18.30 (2.07)	17.83 (2.52)	18.08 (2.18)
School connection	17.24 (2.98)	17.45 (2.99)	17.23 (2.83)	16.99 (3.06)
Note. The means and SDs presented in this table have been calculated using all data present for a particular variable (that is, they are *not* based on complete case data), as per EEF guidance.				

Table 11: Means and SDs of pupil outcomes at pre-test (baseline) and post-test for the Y5 AfA target cohort subgroup

	Intervention		Control	
Outcome	Pre-test mean (SD)	Post-test mean (SD)	Pre-test mean (SD)	Post-test mean (SD)
KS1 reading and writing combined	12.76 (3.45)	-	13.12 (3.44)	-
KS2 reading	-	25.04 (10.18)	-	27.78 (10.01)
KS1 mathematics	13.58 (3.45)	-	13.97 (3.41)	-
KS2 mathematics	-	54.34 (26.30)	-	62.32 (25.49)
Self-esteem	11.32 (1.93)	11.48 (1.90)	11.19 (1.98)	11.34 (1.91)
Goals and aspirations	8.05 (1.74)	8.15 (1.85)	7.90 (1.91)	8.03 (1.84)
Family connection	17.50 (2.76)	17.96 (2.27)	17.40 (2.79)	17.92 (2.43)
School connection	16.98 (3.01)	17.56 (2.82)	17.12 (3.00)	17.09 (2.90)
Attendance (AfA target group)	4.62 (9.55)	5.62 (11.26)	5.46 (12.35)	6.10 (12.76)
Note. The means and SDs presented in this table have been calculated using all data present for a particular variable (that is, they are *not* based on complete case data), as per EEF guidance.				

Missing data

The extent of missing data was determined for all outcome measures for the whole Y5 cohort (see Table 12). As the percentage of missing data exceeded the 5% for the KS2 attainment and the resilience-related measures, additional sensitivity models using FIML or MI were conducted to account for missing data for these outcomes The extent of missing data for the Y5 AfA target cohort subgroup is also displayed in Table 13.

Table 12: Extent of missing data on outcome variables for the whole Y5 cohort[21]

	n (missing)	% missing
Reading	5813 (525)	8.28%
Mathematics	5768 (570)	8.99%
Self-esteem	4297 (1757)	32.20%
Goals and aspirations	4277 (1777)	32.52%
Family connection	4045 (2009)	36.18%
School connection	4097 (1957)	35.36%
Attendance (AfA target group)	1332 (42)	3.06%

Missing data was further examined using a two-level (pupils clustered in schools) binomial (logit) MLM in MLwiN (version 2.22). A pupil-level binary variable distinguishing complete and incomplete cases was entered as the outcome measure,[22] FSM eligibility and AfA target group membership were entered as pupil-level explanatory variables, and trial group was entered as a school-level predictor. FSM eligibility (if EverFSM 6; $\boldsymbol{\beta}$ coefficient = 0.352, SE = 0.119, p = 0.003) and AfA target group (if yes; $\boldsymbol{\beta}$ coefficient = -0.480, SE = 0.115, $p < 0.001$) were significant predictors of the likelihood of having a complete case, but trial group was not ($\boldsymbol{\beta}$ coefficient = -0.050, SE = 0.145, p = 0.731).These findings indicated that missing data was associated with our observed values, therefore supporting the use of MI under the assumption of data missing at random .

Results

Table 7 provides a summary of the ITT and subgroup analyses. Full complete case models for all multilevel analyses are presented in Appendix I along with the results of additional sensitivity analyses (for example, added co-variates and FIML/MI analyses accounting for missing data).[23] Note that the attainment and attendance data sourced from the NPD is based on the *amended* data set.

[21] Based on complete case data—the presence of pre- and post-trial data for a given outcome measure.
[22] Calculated for the primary outcome measure of reading (that is, a complete case corresponding to the presence of a baseline score for KS1 reading and writing combined and a post-test score for KS2 reading).
[23] Where applicable, that is, only for measures where the extent of missing cases exceeded 5%.

Table 13: ITT and subgroup analyses—pupil level outcomes

	Intervention group		Control group		Effect size		
Outcome	*n* (missing)	Mean (95% CI)	*n* (missing)	Mean (95% CI)	*n* in model (intervention, control)	Hedges *g* (95% CI)	*p*-value
Main effects of intervention—ITT (Model 1.1)							
KS2 reading	2764 (263)	32.03 (31.67, 32.38)	3049 (262)	33.66 (33.34, 34.99)	5813 (2764, 3049)	-0.119 (-0.17, -0.07)	0.008
KS2 mathematics	2762 (265)	73.34 (72.43, 74.25)	3006 (305)	77.60 (76.77, 78.44)	5768 (2762, 3006)	-0.114 (-0.17, -0.06)	0.055
Self-esteem	2127 (900)	11.95 (11.87, 12.02)	2170 (857)	11.92 (11.85, 12.00)	4297 (2127; 2170)	-0.005 (-0.06, 0.06)	0.929
Goals and aspirations	2105 (922)	8.41 (8.34, 8.49)	2172 (855)	8.30 (8.23, 8.36)	4277 (2105; 2172)	0.045 (-0.01, 0.11)	0.353
Family connection	1993 (1034)	18.29 (18.20, 18.38)	2052 (975)	18.11 (18.01, 18.20)	4045 (1993, 2052)	0.061 (0.00, 0.12)	0.248
School connection	2004 (1023)	17.45 (17.33, 17.58)	2093 (934)	17.00 (16.87, 17.13)	4097 (2004, 2093)	0.152 (0.09, 0.21)	0.013
Subgroup effects of intervention—pupils eligible for FSM (Model 2.1)							
KS2 reading	820 (62)	29.47 (28.77, 30.16)	893 (80)	30.89 (30.26, 31.52)	1713 (820, 893)	-0.123 (-0.22, -0.03)	0.034
KS2 mathematics	818 (64)	65.32 (63.55, 67.08)	867 (106)	69.99 (68.37, 71.60)	1685 (818, 867)	-0.169 (-0.26, -0.07)	0.021
Self-esteem	619 (2408)	11.70 (11.55, 11.86)	603 (2424)	11.72 (11.58, 11.87)	1222 (619; 603)	-0.037 (-0.15, 0.08)	0.650
Goals and aspirations	620 (2407)	8.37 (8.22, 8.51)	603 (2424)	8.25 (8.11, 8.39)	1223 (620; 603)	0.006 (-0.11, 0.12)	0.940
Family connection	588 (2439)	18.09 17.91, 18.27)	574 (2453)	17.95 (17.77, 18.14)	1162 (588; 574)	0.039 (-0.08, 0.15)	0.636
School connection	584 (2443)	17.41 (17.16, 17.67)	561 (2466)	17.16 (16.92, 17.40)	1165 (584; 581)	0.125 (0.01, 0.24)	0.143
Subgroup effects of intervention—pupils in the AfA target group (Model 2.2)							
KS2 reading	594 (76)	25.04 (24.21, 25.86)	637 (67)	27.78 (27.01, 28.56)	1231 (594, 637)	-0.157 (-0.27, -0.05)	0.026
KS2 mathematics	592 (78)	54.15 (52.03, 56.26)	637 (67)	62.26 (60.28, 64.25)	1229 (592, 637)	-0.185 (-0.30, -0.07)	0.017
Self-esteem	460 (2567)	11.46 (11.29, 11.63)	430 (2597)	11.35 (11.18, 11.53)	890 (460; 430)	0.029 (-0.10, 0.16)	0.719
Goals and aspirations	456 (2571)	8.13 (7.95, 8.30)	433 (2594)	8.06 (7.89, 8.23)	889 (456; 433)	0.016 (-0.12, 0.15)	0.811
Family connection	433 (2594)	17.96 (17.75, 18.18)	410 (2617)	17.92 (17.70, 18.15)	843 (433; 410)	-0.029 (-0.16. 0.11)	0.715
School connection	428 (2599)	17.59 (17.34, 17.85)	415 (2612)	17.05 (16.77, 17.32)	843 (428; 415)	0.152 (0.02, 0.29)	0.085
Attendance	648 (22)	5.61 (4.74, 6.48)	684 (20)	6.10 (5.14, 7.05)	1332 (648, 684)	-0.005 (-0.11, 0.10)	0.666

Notes.

1. The values in this table have been calculated based on complete cases only (that is, the presence of both a pre- and post-test score for an individual on a given outcome measure), as per EEF guidance. 2. p-values in this table correspond to the coefficient in a given model, whereas CIs correspond to the standardised intervention effect size.

Primary outcome—reading

The primary analysis of the impact of the AfA intervention versus business as usual was examined on the basis of ITT using multilevel modelling with random intercepts and fixed slopes. The primary outcome measure of interest was reading. Table 10 summarises the means and effect sizes of the intervention and control group for all pupils in the Y5 cohort (complete cases only).

The means of the post-test reading score suggest that the control group scored on average 1.63 marks higher than the AfA intervention group. When modelled to account for prior attainment (pre-test scores of KS1 reading and writing marks combined), a statistically significant difference was found (see Model 1.2 in appendix Table A1), which when standardised resulted in a negative effect size of -0.119 (CI = -0.17 to -0.07). A subsequent sensitivity analysis that included all additional explanatory variables in the model simultaneously produced the same results (see Model 1.2 in Table A1). In addition, there was a significant group effect of intervention when using MI to re-estimate the models to account for missing data (see Models 1.1 and 1.2 in Table A2). In summary, our analyses indicated that business as usual was superior to AfA in improving pupils' reading scores; this finding was not sensitive to changes in how we modelled the data.

Secondary outcome—mathematics

The main analyses revealed no significant effect of the intervention on mathematics (effect size = -0.114, CI = -0.17 to -0.06; see Model 1.1 in Table A7), but a subsequent sensitivity model that included all additional explanatory variables found that the control group had significantly greater post-trial mathematics scores compared to the AfA intervention group (see Model 1.2 in Table A7). Furthermore, a significant intervention effect was also found when using MI to re-estimate the models to account for missing data (see Models 1.1 and 1.2 in Table A8). In summary, our analyses indicated that the maths scores of pupils in AfA and business as usual schools did not differ; however, this finding was sensitive to changes in how we modelled the data. Both of our sensitivity analyses indicated that business as usual was superior to AfA in improving pupils' maths scores.

Secondary outcomes—resilience-related measures

The main MLMs found no significant effects of the intervention on pupils' ratings of self-esteem (effect size = -0.005, CI = -0.06 to 0.06), goals and aspirations (effect size = 0.045, CI = -0.01 to 0.11), or family connectedness (effect size = 0.061, CI = 0.00 to 0.12) (see Model 1.1 in Table A13, Table A19, and Table A25). These findings were also not sensitive to changes in the modelling parameters—no intervention effects were found on these measures when the additional explanatory variables were entered simultaneously (see Model 1.2 of Table A13, Table A19, and Table A25) or when FIML was used to re-estimate models to account for missing data (see Model 1.1 and 1.2 of Table A14, Table A20, and Table A26). In contrast, the main MLM estimated for the resilience-related outcome of school connection yielded a statistically significant intervention effect (effect size = 0.152; CI 0.09 to 0.21), whereby pupils in the intervention schools had higher ratings of school connection relative to pupils in the control schools (see Model 1.1 in Table A31). Subsequent models that included all of the additional explanatory variables simultaneously (Model 1.2 in Table A31) and used FIML to account for missing data (Model 1.1 and 1.2 of Table A32) produced the same conclusions. In summary, our analyses indicated that most of the resilience-related outcomes of pupils in AfA and business as usual schools did not differ; these findings were not sensitive to changes in how we modelled the data. The only exception to this was that AfA had a positive impact on pupils' school connection.

Secondary outcome—attendance

For the final secondary outcome measure of attendance, we conducted a multilevel negative binomial regression using count data of the AfA target cohort only. There was no significant impact of the AfA intervention on pupil's attendance (effect size = 0.005, CI = -0.11 to 0.10; see Model 1.1 in Appendix J). Similarly, there was no significant impact in the subsequent sensitivity model which included all the additional covariates (see Model 1.2 in Appendix J). In summary, our analyses indicated that the attendance of pupils in AfA and business as usual schools did not differ; this finding was not sensitive to changes in how we modelled the data.

Subgroup analyses

Two sets of subgroup analyses were conducted for each outcome measure by re-estimating the primary models with only pupils eligible for FSM and then again for only pupils in the AfA target group.

In terms of attainment, the results for both the FSM subgroup and the AfA target subgroup mirrored the findings of the main analysis for the primary outcome of reading with the control group performing significantly better at post-test relative to the AfA attainment group (FSM subgroup effect size = -0.123, CI = -0.22 to -0.03; AfA target subgroup effect size = -0.157, CI = -0.27 to -0.03). See Table A3 and Table A4 for a summary of these models. Statistically significant intervention effects were also found when re-estimating models using MI to account for missing data (see Models 2.1 and 2.2 in Table A5 and Table A6). For the attainment measure of mathematics, in contrast with the findings of the main analyses conducted with the whole Y5 cohort, the results indicated that the control group scored significantly higher at outcome compared to the AfA intervention group for both the FSM subgroup (effect size = -0.169, CI = -0.26 to -0.07) and the AfA target cohort subgroup (effect size = -0.185, CI -0.30 to -0.07). See Models 2.1 and 2.2 in Table A9 and Table A10 for a summary of these models. When re-estimating these models with MI to account for missing data, the same results were found (see Models 2.1 and 2.2 in Table A11 and Table A12).

For the resilience-related outcome measures, there were no significant intervention effects for the two subgroups in terms of pupils' ratings of self-esteem (FSM subgroup effect size = -0.037, CI = -0.15 to 0.08; AfA target subgroup effect size = 0.029, CI = -0.10 to 0.16), goals and aspirations (FSM subgroup effect size = 0.006, CI = -0.11 to 0.12; AfA target subgroup effect size = 0.016, CI = -0.12 to 0.15), family connectedness (FSM subgroup effect size = 0.039, CI = -0.08 to 0.15; AfA target subgroup effect size = -0.029; CI = -0.16 to 0.11), or school connectedness (FSM subgroup effect size = 0.125, CI = 0.01 to 0.24; AfA target subgroup effect size = 0.152, CI = 0.02 to 0.29). For a summary of these model results, see Models 2.1 and 2.2 in Table A15, Table A16, Table A21, Table A22, Table A27, Table A28, Table A33, and Table A34. These models were not sensitive to changes made to the modelling parameters when re-estimating using FIML to account for missing data, with the following exception: a statistically significant positive intervention effect was found for ratings of self-esteem when estimated for the subgroup of FSM eligible students only. See Models 2.1 and 2.2 in Table A17, Table A18, Table A23, Table A24, Table A29, Table A30, Table A35, and Table A36 for a summary of these results.

Cost

The cost of the AfA programme delivered during the trial differs from the true cost of delivery for schools buying into the programme outside the context of this trial. Delivery costs for the intervention schools in the trial were subsidised by the EEF, meaning participating schools paid a total of £5,000 for the two-year programme (£3,000 in the first year of the trial and £2,000 in the second year). Schools buying into the programme in a 'real world' setting must pay an annual fee of £5,950 for what is typically a two-year programme, resulting in a total cost of £11,900 (see Table 14 for the cost of delivering the AfA programme and see Table 15 for the cumulative three-year costs of the AfA programme). The annual fee covers running costs of the AfA programme, which includes a physical handbook, access to the Bubble (a digital resource), and 12 half-day coaching visits per academic year.[24]

Table 14: Cost of delivering the AfA programme

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
Annual licence fee	Running cost per school	£5,950	(£5,950 x 2) = £11,900	(£11,900/3/354)[25] = £11.21

A prerequisite of the programme is that teachers involved in the programme have access to a computer with internet access in order to access the Bubble digital resource. It is a reasonable expectation for schools to have the means to access digital technologies, therefore this was not considered an additional financial burden. However, teachers in case study schools did note that additional time was required to navigate the Bubble portal and engage with its resources.

Table 15: Cumulative costs of the AfA programme

	Year 1	Year 2	Year 3
The AfA programme	£5,950	£11,900	£11,900

[24] An introduction to the AfA programme and teacher training is typically delivered during the first two coaching sessions.
[25] Government estimates of school size (number of pupils in school, sourced in July 2019) were used to calculate the average number of pupils in the intervention schools as this was not recorded at the start of the trial.

The AfA programme takes place during the normal teaching day and involves significant staffing requirements outside of normal practice. Teachers were required to engage with 12 half-day coaching visits that took place over the course of each academic year of the trial. Furthermore, structured conversations were expected to take place once per term (three per academic year) with the parents of each child in the AfA target group. These sessions lasted up to 60 minutes. Based on the average of 20 pupils per AfA target group per school (inclusive of both Y4 and Y5 cohorts), it is estimated that teachers should have conducted approximately 60 structured conversation sessions per academic year per school. Working on the average two-form entry intake of participating schools, these 60 structured conversations would be divided between four teachers (two each in Y4 and Y5), meaning that each teacher should have conducted 15 structured conversations per year. As well as the time spent conducting the structured conversations, additional time should be spent communicating with parents to initially engage them and then to arrange sessions throughout the year. Additional time is also required in order to plan and document the conversations. As the structured conversations usually take place during the school day, some teachers need to be released from the classroom, meaning the organisation of supply cover incurs additional time and financial demands on participating schools.

Implementation and process evaluation

Usual practice and control group activity

As part of the usual practice survey (UPS), schools were asked about the type(s) of CPD they had received. We report here on three items that reflect *external*, 'AfA-like' CPD—'coaching', 'accredited training', and 'online learning'; and one—'within-school training'[26]—that mirrors the *internal* process of AfA (whereby an AfA coach worked closely with a school's AfA lead, who in turn worked with internal staff). For each trial arm, we used baseline data (DP1) to calculate the proportion of schools that engaged with these types of 'AfA-like' CPD.[27] Comparable levels of engagement were found between the intervention ('coaching' = 44%, 'accredited training' = 53%, 'online learning' = 9%, and 'within-school training' = 71%) and control schools ('coaching' = 52%, 'accredited training' = 47%, 'online learning' = 24%, and 'within-school training' = 69%) at baseline (DP1). The key exception to this was 'online learning', which was more prevalent in schools that ultimately went on to join the control arm of the trial (approximately one in four schools compared to one in ten among schools that were allocated to the intervention arm).

Table 16: Levels of engagement with external and internal 'AfA-like' CPD at DP1 and DP2

	Intervention		Control	
	DP1	DP2	DP1	DP2
Coaching	44%	44%	52%	56%
Accredited training	53%	49%	47%	56%
Online learning	9%	22%	24%	24%
Within-school training	71%	58%	69%	62%

Note. At DP1, all 66 invention and 68 control schools completed the UPS survey. At DP2, 55 intervention and 55 control schools completed it.

In another part of the survey, schools were asked about the nature and extent of CPD. This section was divided into four categories that mirror the content of AfA: (1) leadership for inclusion, (2) teaching and learning, (3) wider outcomes and opportunities, and (4) parent and carer engagement. For these four categories teachers were asked 26 questions in total. For each, teachers were asked if they had been involved with a certain type of CPD with a binary response option of 'yes' or 'no'. This data was used in an exploratory factor analysis to derive a single UPS score that was entered into the MLMs conducted as part of the impact analysis. In addition, if schools selected 'yes' to any of the questions they had the option to (a) rate on a scale of one to five the extensiveness of the CPD (with five being most extensive and one least extensive) and (b) to make additional notes to provide further details. The ratings were used to establish balance at baseline (DP1)—mean scores and SDs were calculated for the intervention and control groups across the four categories and for the total UPS score entered in the MLMs.[28] Effect size (Cohen's *d*) comparisons were small for leadership for inclusion (0.11), teaching and learning (0.21), wider outcomes and opportunities (0.14), parent and carer engagement (0.15), and the total UPS score (0.12). See Table A39 in Appendix L for a summary.

Change between the two groups over time was also assessed. Scores across the four CPD categories and the total UPS score from both time points (DP1 and DP2) were compared. A two-by-two mixed-measures analysis of variance (ANOVA), with a within-measures factor of time (DP1 and DP2) and a between-measures factor of group (intervention and control), was conducted separately for each of the five CPD measures (see Appendix M for a summary of these statistical tests and descriptives). The results indicated no group differences over time (that is, no significant interactions) with the exception of the teaching and learning measure where the pattern of responses indicated a decrease in CPD in AfA schools relative to an increase in control schools. Overall, these results suggest there is minimal evidence of compensatory rivalry in the control group. In other words, the impact findings reported earlier are unlikely to be attributable to business as usual schools 'ramping up' their activities in relation to the CPD areas noted above.

[26] Items that were excluded from this section include those relating to workshops/lectures, demonstrations/observations, networking, and practitioner research projects.
[27] To do so we created a binary response of 'yes' the school engaged with this type of CPD, or 'no' they did not. A 'yes' was scored if a school selected a response relating to an external source, that is, 'HEI' (Higher Education Institution'), 'LA' (Local Authority), or 'Consultants'.
[28] The EFA conducted with the 26 items revealed that a single-factor model that included a subset of 14 items best accounted for the data. The 14 items were totalled for each school and entered as a school-level predictor into the MLMs conducted as part of the sensitivity analyses for the impact evaluation.

As part of the UPS, schools were also asked about other named initiatives that they were delivering. Business as usual in the control group consisted of a range of named initiatives including, but not limited to: Success@Arithmetic, Art of Brilliance training, Talk for Writing training, Family Links training, Emotion Coaching training, Power of Reading training, Forest School training, Growth Mindset, the MITA Project (Maximising the Impact of Teaching Assistants), Building Learning Power, the Jigsaw Curriculum, Nurture Groups, Circle Time, ELSA (Emotional Literacy Support Assistant; online resources), Visible Learning, ELKLAN (speech and language training for teachers), and IRIS Connect (a video-based professional learning platform).

Implementation analysis

Fifty-three intervention schools completed the implementation survey from April to July 2018 and were scored on five subcategories relating to their reported progress in implementing different aspects of AfA: leadership for inclusion (M = 1.854, SD = 1.474, possible range = 0–15), teaching and learning (M = 2.144, SD = 1.268, possible range = 0–12), provision for wider outcomes (M = 1.491, SD = 1.336, possible range = 0–12), structured conversations fidelity (M = 1.892, SD = 0.762, possible range = 0–3), and structured conversations dosage (M = 1.822, SD = 0.843, possible range = 0–3). Of note is the fact that, given the nature of the AfA programme, the latter two measures are the only ones which correspond to 'traditional' notions of implementation dimensions such as fidelity and dosage. The items corresponding to them were adapted from a similar survey used in the evaluation of the AfA national pilot (Humphrey and Squires, 2011b).

Five schools withdrew from the AfA programme post-randomisation but prior to delivery of the intervention, and an additional four schools ceased implementation from the end of year one of the trial onwards. Hence, the analyses that follow are based upon responses from the 55 schools that did not discontinue implementation. The nine 'withdrawn' schools are not included because they had ceased (or never begun) implementation at the point at which the implementation survey was administered.

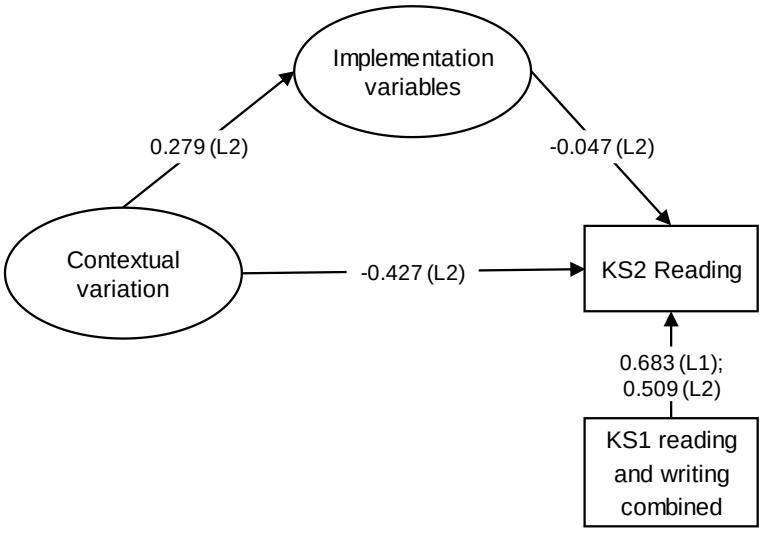
To investigate RQ3b(iii)—the extent to which implementation variability moderates intervention outcomes—two-level multilevel models were estimated with data of the intervention schools only. See Table A37 for a summary of these models. KS2 reading was used as the outcome variable and KS1 reading and writing combined scores were entered as a pupil-level explanatory variable to control for baseline performance. At the school-level, the five implementation variables were entered as predictors: (1) leadership for inclusion, (2) teaching and learning, (3) provision for wider outcomes, (4) structured conversations fidelity, and (5) structured conversations dosage. Results revealed that none of the explanatory variables had a statistically significant association with reading outcomes, with the exception of leadership for inclusion (p = 0.039), whereby higher levels of implementation relating to this component were associated with *lower* reading outcomes ($\boldsymbol{\beta}$ coefficient = -0.726). This finding is counter to our initial predictions and the AfA theory of change as it implies that greater progress in implementation of the leadership for inclusion aspect of AfA was associated with worsening levels of attainment.

Next, two separate models were run to model interactions between the implementation variables and subgroup membership. See Table A38 for a summary of these models. First, FSM was entered as a pupil-level predictor and interactions between FSM and the five implementation variables were modelled, none of which were statistically significant. Second, AfA target group membership was added as a pupil-level covariate and interactions between this subgroup and the implementation variables were modelled, but again, none were significant.

Multilevel SEM was employed to address RQ3b, RQ3c, and RQ4. Path analysis with two levels (pupils clustered in schools) was used to investigate the relationship between contextual variation and the primary outcome, and whether it influenced the implementation of the AfA intervention. The model specification is displayed in Figure 6, along with the standardised estimates at the school level (level two) and pupil level (level one). The measure of KS1 reading and writing scores combined was significantly related to the primary outcome measure of KS2 reading scores at both the school level (p = 0.002) and pupil level (p < 0.001); the magnitude of these associations was, as expected, substantial. After accounting for baseline performance, contextual variation was found to be significantly associated with KS2 reading scores at outcome (p = 0.037). However, the relationship between contextual variation and the implementation variables, and between the implementation variables and KS2 reading, were both non-significant (p = 0.164 and p = 0.784, respectively). Thus, after controlling for prior academic attainment, contextual variation among schools did not

appear to influence their implementation of AfA, and this in turn did not appear to influence KS2 reading outcomes among their pupils.

Figure 6: Model specification—relationships that were examined



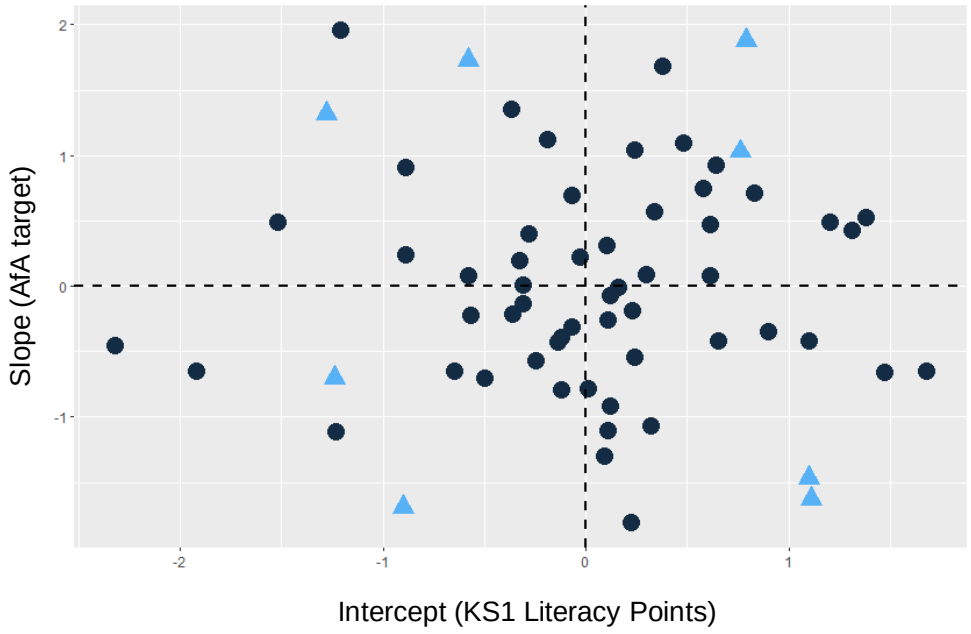
Standardised estimates at the school level (level two, L2) and pupil level (level one, L1) are displayed. The three variables used to derive the level two contextual variation factor were the school-level randomisation parameters—%FSM, %SEN, %RWM 4+ (continuous percentage data was used). The five variables used to derive the level two implementation factor were 'leadership for inclusion', 'teaching and learning', 'provision and wider outcomes', 'structured conversations fidelity', and 'structured conversations dosage'. The primary outcome measure is pupil-level KS2 reading scores. Baseline performance data (KS1 reading and writing combined scores) were also entered into the model.

Case study selection

Intervention schools, *n* = 62—four intervention schools that withdrew post-randomisation but pre-delivery were excluded from this analysis[29]—were ranked according to two dimensions (see Additional Analyses in the Statistical Analysis section for more details) and identified as belonging to one of four corresponding categories: (1) low attaining schools with above average mean prior attainment for the AfA target group, (2) high attaining schools with above average mean prior attainment for the AfA target group, (3) low attaining schools with below average mean prior attainment for the AfA target group, or (4) high attaining schools with below average mean prior attainment for the AfA target group. Figure **7** displays how the intervention schools were distributed among these four quadrants. Schools were approached in rank order within each quadrant until eight schools (two from each quadrant) had agreed to participate in the case study analysis (see the Implementation and Process Evaluation section).

[29] Note that an additional intervention school withdrew from the AfA programme post-randomisation and pre-delivery (in total, five intervention schools withdrew prior to delivery). However, as this school withdrew prior to the case study selection process it was still included in this analysis (it was not selected for case study).

Figure 7: Pairwise residual plot of KS1 literacy points. Low attaining and high attaining schools are located in the left and right quadrants, respectively, and schools with a greater proportion of AfA target pupils with below or above average scores are shown in the bottom and top quadrants, respectively. The eight schools that participated in the case study analysis are shown as light blue triangles. The five intervention schools that withdrew post-randomisation but pre-delivery have been excluded from the plot.



Implementation case studies

The characteristics of the eight case study schools are outlined in Table 17: Case study school characteristics.

Table 17: Case study school characteristics

School	Area	Size	%FSM	Attainment level (compared to average for KS1 Literacy)	Ofsted grade at start of trial
1	London	Two-form	20	Low attaining	Good
2	Kent	Two-form	15	High attaining	Good
3	Greater Manchester	Two-form	20	Low attaining	Good
4	Birmingham	Four-form	29	Low attaining	Outstanding
5	Devon	Two-form	20	Low attaining	Good
6	Lincolnshire	Single-form	16	High attaining	Good
7	Birmingham	Three-form	13	High attaining	Good
8	Bedfordshire	Single-form	20	High attaining	Good

Note. Attainment level is drawn from the residual plot shown in Figure 7.

Case study school profiles

School 1 is an urban community school in London. It is larger than average and is in an area of high deprivation, with approximately 20% of pupils eligible for FSM. The most recent Ofsted grade was 'good'. The majority of pupils attending School 1 achieved below average in terms of KS1 literacy attainment before the trial period. The deputy headteacher was selected to be the school champion for AfA. The school reported previous issues with parental engagement, which is one of the reasons they wanted to be involved in the programme. School 1 ceased engaging with researchers after two school terms, though it did continue to have some communication and engagement with its AfA coach.

School 2 is a rural converted academy school in Kent in a ward with a high level of deprivation. It is larger than average and has a high level of mobility among its families as it is located near an army barracks. The most recent Ofsted grade was 'good' and the majority of pupils achieved above average for KS1 literacy prior to the trial. The school is part of a multi-academy trust alongside two other schools and is a DfE National Support School offering school-to-school support to help develop teaching and learning practices in other schools. The assistant headteacher was selected to act as the

school champion for AfA. School 2 became involved in the programme partly to develop practice for its non-service families, including improving parental engagement among this group, and partly to support speech and language development with its younger year groups.

School 3 is an urban community school in a deprived area of Greater Manchester. It is larger than average and has approximately 20% of pupils eligible for FSM. The school's most recent Ofsted grade was 'good'. The majority of pupils achieved below average for KS1 literacy prior to the start of the trial. School 3 was not connected with other schools upon beginning the programme but was planning to set up these links. The AfA champion role was shared between the headteacher and the SENCo. The school signed up to the programme because the LA suggested that having a school improvement partner would be useful.

School 4 is an urban community school located in Birmingham. It is in an area of high deprivation with almost 30% of pupils eligible for FSM. The school is four-form entry and is more than three times larger than the national average. The most recent Ofsted grade was 'outstanding'. Prior to the trial, the majority of pupils were achieving below average for KS1 literacy. School 4 is identified as a training school and is part of a local school alliance programme. The deputy headteacher took on the role of AfA champion and the school signed up to take part in the trial because it was interested in raising the attainment of vulnerable groups.

School 5 is an urban community school located in a deprived ward of Devon with 20% of pupils eligible for FSM. It is larger than average and the majority of pupils were achieving below average for KS1 literacy before the start of the trial period. At the beginning of the programme, the school had been most recently graded by Ofsted as 'good', but during the trial period the school was revisited and graded as 'requires improvement' across all areas. The headteacher was relatively new in post at the start of the programme. The AfA champion was the assistant headteacher and SENCo. School 5 signed up to take part in the AfA trial because it was interested in improving academic attainment. During the course of the trial, the AfA champion was absent from school for one term. Their role was not taken on by anyone else during this time and there was limited activity; as such, the school and coach were discussing with AfA the possibility of adding one additional term at the end of the trial period.

School 6 is an urban Catholic academy school located in Lincolnshire. It is single-form entry and is smaller than the national average. The most recent Ofsted grade was 'good' and the majority of students were achieving above average for KS1 literacy. School 6 is part of a large Catholic multi-academy trust with fifteen other schools. The headteacher was new in the post when the trial began and took on the role of AfA champion. The school signed up for the trial as part of a broader effort to improve outcomes. School 6 was no longer directly delivering discrete AfA activity by the end of the trial, but the headteacher did continue to meet with the AfA coach, using this time to discuss current issues the school was navigating.

School 7 is an urban academy school located in Birmingham. It is in a deprived area at a national level but is among the least deprived areas in Birmingham. The most recent Ofsted grade was 'good' and the majority of pupils were achieving above the national average for KS1 literacy. School 7 is larger than average. It is part of a multi-academy trust with one other school. The AfA champion was the deputy headteacher who worked across both schools within the trust. The champion had previously been aware of AfA through communication with other schools and instigated the school's engagement with the programme.

School 8 is an urban community school located in a deprived area of Bedfordshire with approximately 20% of pupils eligible for FSM. The school was most recently graded as 'good' by Ofsted and the majority of pupils were achieving above average for KS1 literacy. The school is slightly smaller than the national average and is linked with another school where the headteacher provides support for two days a week. The AfA champion was a deputy headteacher and classroom teacher. The school was interested in taking part in AfA because it wanted to improve parental engagement and academic attainment.

Qualitative data analysis method

See Table 18: Thematic framework for the qualitative implementation and process evaluation of Achievement for All for an overview of the final thematic framework for the qualitative strand of the IPE. Quotes/excerpts provided in the sections that follow were selected on the basis of being particularly illustrative of the themes and codes being discussed.

Table 18: Thematic framework for the qualitative implementation and process evaluation of Achievement for All

Theme		
Superordinate	Subordinate	Major codes (where applicable)
Implementation	Fidelity and adaptation	• Coaching visits • Parent engagement • Teaching and learning • Leadership • Wider outcomes • Online learning portal and resources • Wider networks
	Dosage	• Coaching visits • Structured conversations • Wider activity
	Reach and responsiveness	• Senior leadership team and middle leaders • Classroom teachers • Parent engagement • Pupils and target group • Wider reach
	Quality	
	Programme differentiation	
Factors affecting implementation	Programme systems and characteristics	• Coaching model • Training • Data tracking • Online learning portal and resources • Structure and flexibility • Programme demands
	School factors	• Champion • Senior leadership team • School characteristics • Profile of AfA within the school • School responsiveness and capacity for change • Understanding and expectations
	Teacher factors	• Teacher buy-in and capacity for change • Skills and self-efficacy • Teacher workload and priorities
Perceptions of impact	Leadership and overall school practice	
	Individual teacher practice	
	Parent engagement	
	Pupil impact	
Sustainability	Continuation of programme	
	Continuation of specific practices	
	Factors affecting sustainability	

Implementation

Fidelity and adaptation

Achievement for All is tailored to meet the needs of individual settings, meaning the specific ways that schools engaged with the programme differed. However, there are certain components that school champions and coaches agreed were expected and/or encouraged: schools were expected to engage in coaching visits, to complete a needs analysis and ongoing reviews, to deliver structured conversations (mandatory), and to work on areas of each of the four core modules as tailored to their school. There were also wider opportunities for activity, including the AfA online learning portal and its resources, the Bubble, and engagement in wider AfA networks (for example, other schools implementing the programme).

Coaching visits

Champions and coaches reported that schools generally engaged in coaching visits. On occasion, schools cancelled or delayed visits or disengaged for parts of the programme: 'So that got cancelled […] I never ever got a reply for that in terms of a new date to rebook' (coach, school 4). Initial visits were used to deliver training to teachers and staff members, explaining the programme and the structured conversation approach. Coaches also spent time in the early stages with school champions and members of the senior leadership team (SLT) and took learning walks to understand the school context and priorities. Initial priorities were jointly identified through discussion and were used to shape the needs analysis and inform the focus of later coaching visits; for instance, coaches might go on to review a particular aspect of practice on a visit or spend time with subject coordinators: 'It depends what their priorities are, have some coaching conversations with them or if they want to increase the capacity of middle leaders, say the literacy coordinator, I might have a coaching conversation with the literacy leader' (coach, school 3). These plans often evolved over time based on changing school priorities or wider factors such as Ofsted inspections and exam results: 'Then if you have Ofsted coming in, the key issues might change according to the inspection report and we'll always support with that really. So we're very directed by what the school's needs are, we're very flexible' (coach, school 5). However, at times coaches felt that schools were not deploying them in the best way and so attempted to redirect the focus:

> *'They've already got a very secure plan in place for developing reading. So, is that the best way to use me? So, we were talking about how they perceived how I could support them. So, we've come to an agreement about what we will then do next'* (coach, school 2).

Parent engagement

All schools initially set out delivering structured conversations with parents of target pupils, a mandatory component of AfA. Most continued with these conversations throughout the trial and generally described using the format advised by their coach. However, in almost all cases, the champions reported making small changes to 'tweak' these conversations after the first term, such as developing their own proforma documents to guide meetings, or taking a more structured approach:

> *'Last time it was a bit more wishy-washy if I'm honest […] we've focused it very tightly on reading and only having two targets per child, but two quite tight targets. So that's been a bit more structured this time'* (champion, school 7).

Teachers often further adapted the conversation to fit the needs of particular pupils: 'They picked things out like the fact that some of the SEN children like being creative, so then from that then we think, "Ah okay then we'll build a little bit more around that side of things"' (champion, school 2).

Champions frequently reported that structured conversations did not always happen consistently, often due to the associated time and logistical demands it took to deliver them or the absence of a class teacher: 'The fact that the class teacher left, we had a supply teacher, she left, the structured conversations haven't happened in the way that the scheme requires them to do' (champion, school 6). Over time, various schools gradually withdrew from the component and stopped delivering the structured conversations: 'We haven't done any structured conversations at all because we just didn't have the capacity […] we're not going to be running structured conversations' (champion, school 4).

Occasionally, champions adopted alternative ways to continue engagement with these parents, such as by having regular check-ins: 'I've kept a close contact with the parents of those three children' (champion, school 6).

Teaching and learning

The particular focus on teaching and learning activities taken differed in each case and evolved throughout the trial; several schools adopted a focus on literacy, while other foci included developing the use of teaching assistants, examining marking feedback processes, resilience in learning, and pupils' target-setting. This activity often shifted over time: 'What we've done since Christmas is focus on the resilience [… literacy] was kind of focused on last year' (champion, school 2). Several schools reported monitoring the progress of pupils more closely and others liaised with coaches to examine the particular barriers and support, including interventions, for target pupils:

> *'We're just going to look at what sort of support they get within the class, whether it's TA support, whether it's buddy support, where they're sat, who they're sat with on the carpet, who their talk partners are, when they're on a table who they're with and how do they know what to do when they're stuck? You know, do they know their next steps, their targets?'* (coach, school 5).

Leadership

Parent engagement and teaching and learning activity were the dominant foci in most of the case study schools, with less activity reported for leadership and wider outcomes modules. Coaches often spent time with middle leaders in their capacity as subject coordinators: 'I met with the [literacy and mathematics] coordinators to help them with the priorities that they're developing' (coach, school 6). Coaches occasionally delivered training to the SLT on using coaching with staff, often around monitoring and evaluation feedback: 'He has taken the grow model and he's adapted that into a feedback sheet that all team leaders have used and each team leader then feeds it back to the senior leadership team' (coach, school 3). As in this last example, this support was often seen as a starting point, where the SLT could adapt what they learnt to fit the school's systems and approaches.

Wider outcomes

In terms of wider outcomes, the typical focus was attendance, usually at the direction of the coach: 'Today coming out was that one of the reasons why a couple of children weren't doing well was because of attendance' (coach, school 3). Wider outcomes were often reviewed through termly data monitoring meetings with the coach who provided advice on how to further examine and support this: 'I'm now going to look at the specific strategies that they use to evaluate that […] we're drilled down into particular strategies that have worked well or that need addressing to help the target group' (coach, school 3). Case study schools reported using structured conversations to address wider outcomes, including working with parents towards improving attendance and identifying needs beyond learning:

> '*We noticed that the children whose levels were lower had low attendance, so they had other targets in terms of attendance, so we spoke to the parents about, you know, bringing their children to school'* (teacher, school 7).

Towards the end of the trial period, coaches began encouraging schools to engage with a new wellbeing module that AfA had begun to offer, though there were no instances where champions or teachers reported making use of this.

Online learning portal and resources

Champions and teachers often reported having been shown the Bubble but there were very few instances of staff actually then interacting with this resource, with many staff commenting 'we haven't engaged with that at all' (champion, school 4). As a result, most interaction with the Bubble occurred when coaches used it to facilitate discussion during visits; coaches would then identify and download relevant resources from the Bubble and share these with the champion or teacher: 'Our kind of system now is that I tell [the coach] what I need and she very kindly finds it all and then just sends it to me' (champion, school 5). There were instances where coaches also shared information from wider non-AfA sources, often evidence reviews or evaluation reports relating to the school's focus:

'*Only if it will back the AfA programme or what they need and, you know, you clearly say that's not AfA but, you know, if you feel something is going to be useful to them and it's already out there I think it's not, and you know that, I think it's not nice to not say it*' (coach, school 1).

Wider networks

In some schools, coaches had supported teachers in accessing wider networks across AfA, including linking up with teachers in other AfA schools to learn from their approaches and attending AfA events. For instance, one coach noted: 'We linked up with one school down the road […] their literacy leader came up here and spoke to him about what they're doing' (coach, school 2), while a champion in another school reflected on learning from the conference 'that was quite good, we got like a leaflet about sort of how to spend our pupil premium money and stuff like that, so we can reflect on that yeah' (champion, school 4). However, other schools were not aware of wider networks or events taking place, or were unable to attend due to issues such as distance.

Dosage

Coaching visits

The AfA programme is designed to offer four coaching visits in each school term—a total of twelve visits across the academic year. Most coaches in the case study schools reported that their school engaged in all coaching sessions over the academic year, but that the distribution of these might fluctuate. Some offered more visits in the first term:

'I usually try and get five visits in the first term because of the training so what will happen is, in one term in the future there'll only be three visits, once they're up and running it's fine' (coach, school 6).

Visits focused on conducting termly reviews often had to be delayed until the following term; although these are typically planned for the end of each term, most schools found they could not collate and prepare data from the current term in time for this meeting: 'Often they've done their assessments but they sometimes haven't analysed the data' (coach, school 5). In some cases, schools had fewer visits in one term as they had to cancel due to wider workload or staff absence. Coaches often tried to rearrange any cancelled visits for the following term so that schools still achieved the twelve yearly sessions: 'It averages out at four a term […] even if a visit's cancelled if we are able to rearrange it we will do that' (coach, school 6).

Structured conversations

Structured conversations with parents of target group pupils were expected to take place once every term. While numerous case study schools reported keeping up with the practice for most of the trial period, others struggled to maintain this regularity. At times the school's wider workload made offering one per parent per term unrealistic given the logistics and time that structured conversations required: 'We didn't have the expectation that they'll have structured conversations three times, so we knew it was going to be two realistically knowing how full the school diaries are' (coach, school 1). The structured conversations were not always completed in one term and so the subsequent term became more focused on completing the remaining sessions, sometimes because they had started later than intended:

'I think what's been really frustrating for me are the structured conversations [from the first term] not having finished by [the second term], I've found that really frustrating because I wanted it to start from February and I did a letter but I can't do anything about that' (champion, school 1).

Wider activity

As wider activity for AfA was flexible and agreed between the coaches and champions, it is difficult to comment on the frequency of delivery beyond coaching visits and structured conversations, which are the only components with specified frequency. However, there were some comments around the regularity of wider activity. Some champions and teachers commented that they would have one coaching visit on a particular aspect and then the focus would shift, so that this was never directly revisited or followed up. Some schools commented on this in relation to the leadership module, as the SLT or middle leaders would engage in a coaching visit and then not speak with the coach again or review any changes they made as a result: 'She did one session which included coaching about coaching but there was never any follow up, there was never any opportunity to monitor how it was going' (champion, school 5).

Reach and responsiveness

Senior leadership team and middle leaders

In initial visits, coaches mostly interacted with school champions as they developed their understanding of the school and created an initial needs analysis and action plan. In some cases, this early interaction also extended to members of the SLT beyond the champion, such as the headteacher (if not the champion). Over time some schools saw more direct engagement with the SLT, often focused on roles and responsibilities that overlapped with the school's needs analysis:

> *'That was one of the things that we did identify initially on the needs analysis was intervention groups and [the coach] did come out and talk to our assistant headteacher about interventions, because we didn't quite know where we were going with that'* (champion, school 8).

In some cases, as noted previously, coaching sessions were delivered to members of the SLT around their leadership practice. However, in some schools, coaches and champions felt that the wider SLT staff members had been less directly engaged in the programme than they would have liked, with interaction and activity more limited to the champion: '[The headteacher] did say he would join us this morning at some point […] and it's like, "Well, where are you?"' (champion, school 5). Engagement beyond the SLT was often limited to subject coordinators and SENCos to address particular aspects of their roles and support wider practice. SENCos were often seen as an important person to involve given that target pupils were frequently also engaged in wider SEND provision: '[Meeting with the SENCo] was a really productive session and a lot of the children who are in our AfA group are also SEND' (champion, school 7).

Classroom teachers

In all case study schools, teachers attended training with coaches in the early stages of the trial period. Following this, interaction with the coaches and the programme beyond this point was generally limited: 'Only the teachers who are actually doing the direct work with the structured conversations [are involved with AfA]' (coach, school 6). Several teachers could not identify how AfA was related to their practice beyond structured conversations: 'I don't feel I could identify daily a specific area where I think, "Oh, I've got my AfA"' (teacher, school 2). Some felt that only the teachers delivering structured conversations were aware the programme was still in place in the later stages of the trial:

> *'Unless they're doing the structured conversations I wonder how much they are aware of what we're doing and how we're feeding into the project […] only pockets of staff have been involved in Years 5 and Years 6; I think everybody else has probably forgotten about it if I'm completely honest'* (SENCo, school 7).

Coaches also reported delivering training and advice to schools regarding their use of teaching assistants, but generally did not report spending any time with them as part of this process. In some cases it was felt that there might be changes to practice that teachers and teaching assistants were implementing but that they would not necessarily know this was linked to AfA: 'They're doing a number of children who've been identified for certain interventions, they probably haven't realised it's come from AfA' (champion, school 7). It seemed generally that this was how coaches sought to operate, working predominantly with champions and those leading specific areas of practice with the aim of 'almost indirectly' (champion, school 6) influencing classroom practice.

Parent engagement

Given that structured conversations are a key component of the AfA programme, all schools reported some level of parent engagement. Generally, schools reported that most of the parents they invited did then attend: '[Our school] had a high percentage of parents coming in which is very positive' (coach, school 3). However, it was frequently highlighted that parents initially thought something must be 'wrong' when they were first asked to come in for this meeting, which could be a barrier for engagement:

> *'The biggest challenge I've come across—and it's not a great one—is that people automatically go onto the defensive, so they think there's something up with their child because they've been selected. So, "Oh there must be something wrong because my child has been picked", not the opposite of that which is, 'We feel we could do so much more for your child'"* (teacher, school 2).

Some schools encountered particular challenges in engaging parents for these first meetings, with some being very difficult to reach and actually meet with. Champions occasionally found it frustrating that this was a parent engagement exercise that started off with a need for an engaged parent; that is, if the school had existing issues with engagement, they were ill-equipped to get hard-to-reach parents into the school to even begin the conversation:

> *'Those are parents that are really difficult to reach. I know that's the whole point but if you can't get them through the door it doesn't matter how well trained we are because we can't get them through the door'* (teacher, school 4).

Indeed, in several schools there were some parents who never engaged with the conversations. Overall, however, champions and teachers felt that parents who did engage responded well to structured conversations and reported that most continued to engage throughout the course of the trial, including those who were initially hard-to-reach:

> *'There were some hard-to-reach parents who teachers had tried several times to reach and she was encouraging them to keep trying and actually supporting them with that contact as well and I think she was really positive and so were the staff, really positive about those relationships they'd made with parents'* (coach, school 5).

Pupils and target group

Overall, pupils were not thought to be directly exposed to any component of the programme. Instead, it was felt that AfA would impact child outcomes indirectly through changes to practice: 'It will have an indirect impact through changing sort of staff perceptions and staff ways of working really' (SENCo, school 7). There was usually some level of focus on target pupils, though this differed across each case. In some schools there was a strong emphasis on developing practice and reviewing the strategies used with target pupils, whereas in others this was incidental or a small component: 'We're not just obviously using it on those children because all children are in the class of another twenty odd children who are also having the same' (teacher, school 2). Sometimes schools felt that the status of being an AfA target pupil might simply mean that there was greater day-to-day awareness and attention given to individualised needs during planning: 'In that meeting you know teachers will say, "Well, they're an AfA child", so I think the children who are AfA are very much on everyone's radar' (champion, school 3).

Schools used slightly different approaches to selecting who was in their target group, and at times this influenced their later approaches. For instance, while some selected pupils were identified as the lowest achievers, others further refined this decision based on their specific context, such as selecting pupils likely to stay in the school during the programme in an area with high mobility: 'We felt, really, because it's pointless picking a child for a programme when they might not be here after, half way through it' (champion, school 2). In several cases many of the target pupils were also identified as having SEND and there were times when schools then had to negotiate the additional practices in place and how this would fit alongside activity for target pupils. On occasion they felt uncomfortable implementing additional practice for only some students or were reluctant to include structured conversations for parents when they already had such close contact with these pupils and their families:

> *'I haven't put interventions [in place] because my children are SEND. It hasn't been those children, it's been all of my SEND children, and I cannot prioritise a particular group of SEND children over another other than in terms of need'* (SENCo, school 4).

Wider reach

There were occasions where elements of the programme continued beyond the immediate school context to broader domains. This could involve wider school stakeholders, such as governors; in some schools the SLT were sharing activity and learning with governors to discuss next steps, and in some cases the coaches met with the governors to share information about the programme and inform them of future goals. For instance:

> *'I think governors have been fed-back to; I think it's given them a much clearer idea about our pupil premium and about some of the barriers and some of the things in school we need to tweak. So, I think they've benefitted the way I have from spotting some of the issues and obviously they then monitor that'* (champion, school 7).

Some of the case study schools were part of clusters or trusts where staff worked across different schools. As such, some of the changes in practice were then rolled out across several linked schools: 'The cluster has now adopted those things [attendance policy changes] as well so there's six or seven schools that have actually benefited from that, so that's definitely altered practice' (champion, school 3). In these schools this also included some staff managing split roles across multiple settings, including members of the SLT and the SENCo. As such, while the coach may not visit those schools and so direct activity would not be taking place, staff might develop their practice in smaller ways and then carry this across the schools:

> *'You know, we're talking about leadership for inclusion, so it's encroaching on leadership. So I'm not sending him off to the other school with a lovely pack of information that he can take away, or I'm just letting him [the SENCo] think about his leadership'* (champion, school 3).

Quality

A number of approaches were used by coaches and champions to strengthen the use of the programme. Many of the coaches spent time attempting to gain an in-depth understanding about the particular needs of a given school and attempted to ensure that they were appropriately tailoring the programme to fit these: 'We make an initial visit and we get to know the people who are involved, we chat to the head, we meet the school champion […], we're very much led by the school' (coach, school 5). Several made accommodations in an effort to ensure structured conversations provided 'quality time', such as releasing staff from other responsibilities during the day and ensuring they had planning time:

> '*It's no good planning a meeting at the end of the day when they want to go home and cook dinner or the teachers have had a long day. It needs to be at a time that's suitable for both; yes, it will mean harder work for us in order to manage that time, but it has to be a suitable time for the parents and quality time I would say'* (champion, school 1).

Coaches and champions frequently described checking in with members of staff to review various elements of activity and different approaches being used to identify areas for improvement and future directions, most often in relation to structured conversations:

> '*Obviously individual needs come out of that and that's important, but is there anything strategically that we need to be doing with parents? […] We need to meet as a group, I think, and talk about what's come out of those conversations and how we move forward with it'* (coach, school 2).

Programme differentiation

Champions and teachers frequently identified structured conversations as something that was distinguishable as a distinct practice, and on several occasions they commented that this was more useful than the use of previous approaches like parents' evenings because there was greater opportunity to gain new information and build relationships: 'When we have parents' evening it's like ten minutes, well that's nothing is it? […] I think [structured conversations] is just giving us a level of information that class teachers haven't had before' (champion, school 5). In several cases schools felt that these conversations overlapped so much with existing parental engagement approaches that they had chosen to reshuffle practices, ending parents' evenings for these families or building in Individual Education Plan (IEP) reviews as part of the conversations. Champions and teachers felt that AfA offered a number of further characteristics that set it apart from previous approaches, including the coaching model, the tailored approach, the long-term nature of the programme, and the focus on target children:

> *'I think what's been good about it … it's been real, it's been about our children, so when you go to courses it's about, you know, something you can apply to your children but I think the conversations here have been about our children'* (champion, school 2).

Schools reflected on how the approach built on existing practices and worked closely with their school priorities, 'dovetailing' into the work they were already doing. However, this also meant that schools consistently struggled to identify clear areas of practice that they could attribute definitively to AfA, with the exception of structured conversations: 'There's not really been anything concrete to it' (champion, school 3).

Factors affecting implementation

Programme systems and characteristics

Coaching model

Coaches were typically seen as an important component of the programme. They were generally viewed as experienced and knowledgeable, with champions frequently highlighting their previous school leadership experience: 'Knowing [a coach's] background as well as being a head in the past and stuff so I think that's useful because I think that that then gives her then quite a lot of credibility' (champion, school 2). However, in one case the school felt that there was a misalignment between the coach's background and their own context, which they found particularly challenging: 'We're coming from different perspectives and where her perspective was more for preparing them for [secondary school outcomes] […] mine is right here in primary school' (champion, school 8). Champions talked about valuing the coach as a 'fresh pair of eyes' (champion, school 4) who could offer an external view of the school without being associated with any kind of regulatory body. Several champions commented that coaching visits kept them on track with delivery of the programme, reflecting that:

> *'Having somebody coming in to see you does tend to mean that you do things as well that gets, again with the "initiative madness" that we have in school sometimes […] I know that [the coach is] going to be coming to see me so I need to make sure that I'm doing it'* (champion, school 2).

Several champions used their time with the coach to reflect on their general practice and identify ways to move forward, which they felt they did not always have the space and time to do:

> *'It's having the time really to step back and to really concentrate your thinking and we never have time to do that, do we? […] rather than just firefighting all the time, that you have actually, you know, sort of really reflected about things'* (champion, school 5).

Training

The initial training sessions delivered by coaches on AfA and on structured conversations were generally seen as a useful step in engaging staff: 'I think the training, you know, reiterated the importance of that and gave them particular skills' (coach, school 3). Several champions and teachers commented that this was distinct from other training they had been given in that it was tailored to their school, rather than the 'blanket' approach typically applied in training sessions. However, some questioned why all staff needed to attend these sessions when only some teachers would be actively engaging in AfA activity: 'A good half of the room were never being asked to do those conversations and therefore they were automatically switched off' (SENCo, school 4).

Data tracking

The data tracking that coaches and champions engaged in together for termly reviews often happened later than planned, as highlighted previously, because schools had to prepare the data for these sessions and were not always able to do so in time. Some champions considered the evaluation approaches used by coaches were overly simplistic compared to their own systems, as they failed to provide in-depth or nuanced information about how target children were progressing:

> *'We don't think it's particularly helpful because it says something like "below expected", "expected" or "above expected" […] I don't think that can show in a progress way, I think perhaps if there's be more liaison with how schools record data we could have shown more progress'* (champion, school 4).

Several champions and teachers highlighted that it was incongruent with their existing systems, which introduced another task for them to manage: 'It was a data burden because we've got our own in-house tracking' (champion, school 4).

Online learning portal and resources

As noted previously, schools typically did not make use of the Bubble and its resources. There was a general agreement that this *could* be a valuable resource, but champions and teachers felt it required a substantial amount of time that they did not have:

> *'I think [reviewing the Bubble with the coach] was quite useful… but then, when do you find the time to do it [use the Bubble materials] and actually get the value from it?'* (champion, school 6).

They reported that the volume of information and resources available on this portal could be overwhelming and difficult to manage, and the interface was commonly highlighted as non-user-friendly: 'It has a wealth of information on there, which is a good thing, but I would say almost overwhelming […] I mean, just navigating it—there's so much on there' (SLT teacher, school 8). Though the system was updated partway through the trial, with coaches commenting that it should be easier, qualitative data did not indicate a general uptake in this resource, though one champion commented: 'It's better since it's been updated. I probably looked at it, you know, two or three times a term' (champion, school 3).

Structure and flexibility

There were mixed perspectives on the level of structure and flexibility within AfA. The flexible approach was seen as positive because it enabled a tailored approach and let schools pursue new directions as they arose: 'I think it's probably better […] it's kind of a negotiated process' (champion, school 3). However, at other times this flexibility was viewed negatively, as staff struggled to grasp what the programme was and felt they did not have the time or capacity to make sense of how they should move forward given the volume of available options. Various staff members found it difficult to pinpoint and articulate exactly what the programme was as they felt it was unstructured and 'airy fairy' (champion, school 8). As such, some felt that a slightly more prescriptive approach could be useful: 'I think maybe there needed to be a closer needs analysis on the data and then very much more of a prescriptive approach rather than kind of a lot of general discussion [… with] definitive timescales.' (headteacher, school 5).

Programme demands

Champions and teachers felt that certain aspects of the programme came with a number of time, logistical, and financial demands. As highlighted already, the Bubble was seen as challenging to engage with because it required the time to navigate the online portal and work through the resources. Structured conversations were also seen as a particular challenge. Initial communication to engage parents took a great deal of time in some schools, particularly with hard-to-reach parents: 'We've probably wasted a bit of time on this occasion trying to chase parents' (champion, school 8). Delivering up to an hour-long conversation with a large number of parents every term, often during the school day, was difficult to facilitate. The SLT needed to release teachers from the classroom as well as ensuring that there was sufficient time available for teachers to plan and document the conversations: 'Structured conversations take a lot of time and organising cover and organising supply cover' (champion, school 3). This also meant that they often had to spend a large amount of money on supply teachers: 'You're talking about maybe four, five hundred pounds a time, sixteen teachers, you know, you're looking at eight thousand pounds to do that' (champion, school 3).

School factors

Champion

The school champion was viewed by both coaches and schools as a key component within the programme. In most schools the champion was the main point of contact for the coach, acting as the gatekeeper between the programme and others within the school: 'I know that [the champion] feeds back regularly both to [the headteacher] and also to the others in the senior leadership team' (coach, school 3). The champion typically led and directed AfA activity within the school and was seen as the main driver of the programme, keeping staff on track and motivated where they could and reviewing the quality of delivery: 'What [the champion] brings to it is, I suppose, tremendous organisation and leadership skills and that really drives it through; where that's lacking it really makes a difference to a programme' (coach, school 3). Coaches chose to make use of the champion in different ways. Some focused on working very closely with the champion throughout the course of the programme, with the champion then sharing learning across the school, whereas

others transitioned over time to working more closely with the wider school staff, in which case the champion took on a coordinating and monitoring role:

> *'He's very strategic in his field, you know, he's like, "This is where you need to really use the coaching support, this is the person that needs to work with me this time, this is the team of people". So he organises, you know, the managerial side of things'* (coach, school 2).

The particular role of the champion within the school hierarchical structure was seen as important. Coaches and school staff frequently highlighted that when the champion was a member of the school's SLT this provided greater capacity to affect change and lead others: 'We're in positions of influencing, which [means] we can implement change' (champion, school 4). In all case study schools the champion was a part of the SLT team; however, several coaches reflected on challenges they had experienced in previous schools where this was not the case:

> *'If they are not a member of SLT that proves really difficult because then they can't drive things forward because they have to go to someone else to ask them to drive it forward, who then goes to SLT and has a chat about it and it may not happen'* (coach, school 8).

Working with a champion who had been in the school for a long period of time was seen as beneficial due to their additional knowledge regarding the broader school landscape and priorities, as well as their credibility within the school: '[The champion]'s been at the school a long time; she knows the families, the children, the parents inside out. She's got this whole history behind her, and she's got a lot of credibility in the school' (coach, school 5). Where this was not the case, champions often struggled to appropriately gauge how the programme should be used:

> *'As a new headteacher […] I think—had I had the time over again—I wouldn't even be looking at this until eighteen months or two years into being here when I would be much clearer about where the school was and what the systems were underlying it'* (champion, school 6).

In a number of cases there were issues around champions' availability and capacity to undertake the level of time and work required for AfA, often because they were SLT staff navigating a number of competing roles and responsibilities: '[I] just haven't got the time or the capacity to kind of just be facilitating it all the time' (champion, school 4). In instances where the champion was absent for a period of time (or became disengaged), activity generally stopped altogether in the schools and coaches had limited communication with other members of staff:

> *'The statutory aspect of my work was being covered by somebody else, but obviously none of the rest of it was […] you haven't got that person driving it in the same way so of course it's going to have had an impact'* (champion, school 5).

Various schools attempted to implement a split role, with two champions sharing the workload, though in these cases it seemed that over time one person began taking on the bulk of the work and engaging more with the coach: 'It's more [champion 1] that I see. But I always copy [champion 2] into any communication and I know that [champion 1] feeds back regularly' (coach, school 3). In one school the champion left towards the end of the trial period and so a new champion stepped in, which seemed to slow activity and progress down as the new champion needed time to become familiar with the programme and develop their 'ownership':

> *'I think before [the past champion] was driving it but of course [the new champion], being brand new, he's been a bit limited in how he's been able to drive it because, of course, his understanding of the programme hasn't been there'* (coach, school 8).

Senior leadership team

Both champions and coaches felt that when the SLT recognised a need for AfA and offered support and involvement, this facilitated the programme: '[In some schools] you can see that the champion very much is supported by SLT so again you can see it driving' (coach, school 8). Though as noted above, champions were often part of the SLT—they were typically deputy or assistant headteachers rather than headteachers—and so talked about the need to have the headteacher and other members of their team on board with the programme. In several schools the headteachers did not engage with the coach or seek to be involved with the general AfA activity happening within the school, which coaches and champions found frustrating: 'I'm sure he would just say, "Oh yes, but you know I know you will do it,

you're more than capable", but that's not the point is it?' (champion, school 5). Alternatively, there were other occasions when the headteacher oversaw and advised on all of the champion's decisions, which was seen as problematic; in these instances both champions and coaches felt that the champion needed greater freedom to make decisions and lead activity: 'Everything has to go the headteacher and that will cause organisational problems' (coach, school 1). Indeed, several coaches and champions commented on broader elements of professional trust and freedom around the school's existing leadership model: 'He allows them freedom to do what they feel is good within the classroom as long as there is evidence of impact and I think you know you can see the staff value that' (coach, school 3). Very often a key role of the SLT was facilitating champions and teachers in taking on an additional workload, particularly around structured conversations:

> *'He had also come up with a timetable for the structured conversations. And he has dedicated staff meeting time so that they can conduct a conversation in that time and he has gone over and above the number of sessions that the teachers with the targeted group need, so he allowed for planning time as well […] so it is really, really positive how that is developing'* (coach, school 3).

School characteristics

School size was discussed by some case study schools, with difficulties arising for particularly small or large schools. In small schools, the leadership team was typically smaller and so staff were attempting to manage multiple roles, which meant they struggled to fit AfA into their day-to-day priorities:

> *'In a bigger school then I would expect there to be somebody who would be able to or would be tasked with this as their job […] whereas, at the minute I've got this job where I'm trying to fit it in, but I've got that on top of the other myriad of things that I'm doing, so it really struggles in a small school because of a capacity thing'* (champion, school 6).

Conversely, in larger schools the size of the target group was seen as problematic: 'One of the issues we have found is how many children there are on the programme; I think that's been quite overwhelming […] it's just made it a real technical kind of nightmare' (champion, school 7). In these larger schools, some expressed frustration that AfA were asking them to engage 20% of particular classes because they felt they needed greater flexibility on this expectation in order to make the conversations practical: 'I think it's been impractical for our school to try to facilitate 46 structured conversations […] I think possibly it could have been thought through a bit better that they were the challenges of a school our size' (champion, school 4).

Various schools experienced staffing instability during the trial period, which posed a number of issues and often meant that the programme slowed down while the school was attempting to keep general practice running smoothly: 'Those challenges mean that thoughts of AfA can just be pushed to the one side while they're dealing with those so if you have major staffing issues you have to sort that first before you can do anything else' (coach, school 6). In several schools, particularly smaller schools where staff absence could have greater implications, this was an ongoing issue throughout the trial period: 'The progress for AfA has been curtailed, continues to be curtailed, by the staff instability that we've had […] it required there to be a consistent staffing structure' (champion, school 6). Instability in the SLT could also create issues as schools were focused on managing this change: 'They've been really hard to engage this term […] because they're appointing a new headteacher' (coach, school 4).

Profile of AfA within the school

The profile and prioritisation of the AfA programme within the school, both for the SLT and for the wider teaching staff, shifted throughout the trial period. Often schools felt that the initial training sessions were useful for getting both the SLT and teaching staff on board but that their awareness and motivation lessened over time. As a result, some champions and coaches reported taking measures to try to keep the programme on teachers' radar throughout the trial period: 'We just talked about re-launching it a little bit in September. So I'm going to have a [session] with the teachers involved and just really raise the profile of AfA and why we're doing it' (champion, school 7). Indeed, several champions and teachers felt that teachers in the school were not particularly aware of what was happening as part of AfA:

> *'I think there's a bit of a learning curve there, maybe, with us just raising the profile of it in school and making sure the teachers understand. I think as senior leaders you understand why this is going to work and what's*

happening and I don't think we've been—or I don't think I've done as good a job as I could of explaining that to the teachers' (champion, school 7).

School priorities often shifted and changed throughout the course of the programme. As outlined previously, school priorities can drive programme activity: 'Some of the things that have been thrown up from the Ofsted report—some of the actions they need to carry out—I've been able to support them through AfA' (coach, school 5). As explored, this flexibility and ability to adapt to particular needs was often seen by champions and staff as a valuable feature of AfA. However, while the nature of the programme means that these priorities should be built into ongoing AfA plans, in reality, wider school priorities frequently competed with the programme and began to overshadow it. Often this was reactive as schools attempted to respond to Ofsted actions or manage emerging issues:

'The report that we produce independently of AfA actually has raised the question about mind-set and development and progression […] with the scoping audit and going into amber category, we had to shelf some things and other things became more important' (champion, school 8).

These clashes occurred as champions and the SLT refocused their attention on another area of practice. This could slow down the momentum of the programme and lower the general profile and emphasis on AfA across the school. At times coaches gave schools some space, during particularly busy periods, because they were aware that they did not have the capacity to manage the programme: 'The thing is, right now there isn't much point in, you know, pushing things in the last two three weeks [before the Christmas break] because schools are so busy' (coach, school 1).

School responsiveness and capacity for change

The level of responsiveness differed across case study schools and also fluctuated throughout the programme. A number of coaches found their schools to be very responsive to the programme, engaging in coaching visits and ongoing conversation:

'Totally responsive, dates booked in January for the spring focus. Everything, two dates kept, everything went on ahead, meetings went on ahead so very responsive and in fact, you know, teams keep working together to learn more so yeah, […] very responsive, very committed' (coach, school 7).

However, other coaches found that the school champion and the wider SLT were not very responsive, often cancelling visits and failing to communicate between visits: 'It has been a little bit of a slow start with them' (coach, school 1). There were also differences in the extent to which schools were responding to coaches' suggestions and actually making changes between visits. For some, this was consistently taking place:

'They're very, very responsive; everything that I talk about and suggest it is acted on and I'm confident. So, for example, another example of not just having a look at the title of what I send, but I sent a paper on teaching and learning, and what it was talking about was that the head needs to spend less time on administrative tasks and more on the focus of teaching and learning. [The headteacher] had read, he had digested it, he said "yes that's important" and I think he has taken stock and that has influenced what he's doing. I'm very confident that anything I share he won't just not look, he will look and think about, "Is this right for my school?"' (coach, school 3).

In other cases, coaches found that they needed to keep the 'momentum going' (coach, school 2), finding that schools frequently reduced their activity level between visits. As noted above, these fluctuations were often in response to changes in wider school priorities: 'They're not deliberately being unresponsive; it's the school context that's made them not be able to engage at the moment' (coach, school 4).

A number of coaches and champions reflected that the nature of AfA required a school that was open to change, and that when this wasn't the case this should be developed: 'It's developing that culture of openness and a real understanding of change and why it's happening and putting, you know, learning at the heart of everything' (coach, school 5). Indeed, champions and teachers frequently commented that the school's capacity for change as a result of the AfA programme was due to the wider school context, rather than AfA itself: 'We're open to change […] I don't think in this school that we're in a straightjacket, I don't think we're like academies, I think that we do try things. If they don't

work then we'll try something else' (champion, school 1). Some suggested that they were open to changing purely because they had already identified a need to do so:

> *'I think because of where we are—in a position where things have to change, so I think we're completely embracing it because we need to; we know we've got to move forward. Whereas, I guess, in another school that's got everything sewn up they might be less willing to because, like, we were saying then, if you do something then something else has to go'* (teacher, school 5).

Understanding and expectations

Champions and other members of the SLT frequently commented that at the beginning of the programme they had struggled to understand what it would look like in practice: 'I just want it to start when it starts then I don't know what's going to happen, so I'm quite eager for it to start' (champion, school 1). This issue was present for both the SLT and the wider staff body, as teachers were often very unclear at the outset what was going to be happening: 'Initially people were a bit like, "Ugh! What are we doing here, what's the purpose of this?" […] personally, you know, it wasn't made clear as to how it would be used' (teacher, school 8). Schools frequently reported that the programme was not what they had initially expected, often because they expected a prescriptive intervention rather than a 'school improvement tool'. They typically thought that they would be provided with a specific framework, or that someone would come in and work directly within classrooms: 'I thought that they would be coming along and actually working with [the teachers] in there […] actually coming along and it would be like a more sort of hands-on in teaching and learning' (champion, school 4). On occasion, schools felt that this was better than they had initially expected: 'I think I expected it to be more prescriptive, you know, that "here is a model for you to do" […] that doesn't mean what it's actually like I think is any worse, I think it's probably better' (champion, school 3). However, in other cases the response was not as positive and schools felt that the programme did not align with what they had wanted to achieve: 'I think it's been a bit removed. It's not as hands-on as I'd have liked […] I don't really think it's given us any clear direction' (headteacher, school 5). Often this tension was around structured conversations, which a number of schools had not understood was a requirement within the programme:

> *'We didn't know about structured conversations when we signed up to AfA, so we signed up for AfA because it was the four things of, like, leadership, teaching and learning, parental engagement, and aspirations and wider outcomes. So we thought, "Oh yeah that looks great!" We didn't know about the structured conversations […] and so we feel, like, it's been something that's sort of been done to us'* (champion, school 4).

In some cases, champions felt that during the initial sign-up stage, AfA staff should have been clearer about the nature of the programme and the requirements and the difficulties that it encompassed so that they could make a more informed decision:

> *'I think before you sign up for it, it should be made clear that there's a huge commitment in terms of the requirements to complete structured conversations. So I read back, thinking, "Did I know about this?" and I didn't, it said about engagement with parents'* (champion, school 4).

A number of schools became less engaged in the programme over time because it was not delivering the specific types of activities or changes that they had expected to see.

Teacher factors

Teacher buy-in and capacity for change

The extent to which teachers bought into the programme was identified at the outset as critical: 'I think we need to get buy-in from the teachers […] so it's not a tag-on but it is actually a really valued process' (champion, school 7). Both champions and coaches recognised that changes needed to be supported by all staff, rather than just the SLT:

> *'I think perhaps maybe their biggest obstacle might be […] making sure that everybody buys into it and understands it, because in my conversations with [the champion] she has expressed a concern that the changes are being really driven by the headteacher and the deputy head and perhaps not full involvement of the senior leadership team and perhaps not full buy-in from the staff'* (coach, school 5).

Coaches and champions felt that the initial training was useful in getting teachers on board with the programme: 'They all seemed to go away really quite positive about it, I felt […] my understanding is that the majority of staff have taken this on board' (coach, school 6). Several coaches reported that staff were generally aware of who they were and what was happening and were receptive to the activity taking place: 'Very enthusiastic, nice to hear the staff talking about AfA' (coach, school 7). However, some teachers resisted adopting new changes as they felt that their existing practices were better: 'With the Year 6 they're, like, hard; they have their own ways of working and it's hard to breach that' (teacher, school 4). Indeed, as previously outlined, teachers didn't always have a clear understanding of what the programme was or of the wider activity taking place because they were generally only involved in structured conversations: 'I think everybody else has probably forgotten about it if I'm completely honest' (SENCo, school 7).

Skills and self-efficacy

Initially, many schools experienced initial staff resistance in relation to structured conversations as teachers felt that this was not a new practice for them and so were frustrated at having to attend training on parent engagement: 'Their attitude to structured conversations was, "Oh, this is nothing we're learning for the first time, we do this, we do structured conversations this way anyway"' (coach, school 1). However, many found that this was more helpful than they had expected, given that in most cases this was ultimately viewed as a distinct approach:

> *'If I'm going to be really honest with you, when I was told that I had to go and listen to somebody telling me how to have a structured conversation with a parent I thought, "Oh, you know, with the greatest respect, I don't want anybody telling me how to do that, I'm very good at my job." […] After a very short space of time I think you realise that it's not about them looking at you and judging you as a teacher, it's about supporting the process of collaboration between teacher or school and parents and that there's no right or wrong in it.'* (teacher, school 2).

Following the training, however, some teachers became nervous about delivering the conversations: 'I think they built it up for some reason as being quite difficult and actually they weren't. I think it was a case of "actually, yeah, I've done one and actually it was okay"' (champion, school 2). Over time, champions reported that teachers became more confident in this new area of practice, which they felt contributed to overall levels of buy-in:

> *'I think people have probably just become more confident and more, you know, just really positive […] that degree of positivity, I think, has just become more and more because there's no sense of, "Ugh! It's one more thing on my list of a hundred things I've got to do"'* (champion, school 5).

This training and support were seen as particularly useful for newly qualified teachers (NQTs), both for structured conversations and for general practice:

> *'I think [NQTs] benefitted from the structured conversation training … I think there's some good stuff for their teaching as well for their pedagogy and the resilience. So I think they've benefited from it quite well, because being young teachers I think it's been good for them and I think for us as well as a school, I guess it doesn't hurt to have other views'* (champion, school 2).

Similarly, for teachers who had recently taken on new roles and responsibilities, such as SENCo or leadership positions, AfA was seen as a particularly useful resource to aid development within these new roles: 'As a new [subject] coordinator it really helped her to see the way forward and write her action plan' (champion, school 8).

Teacher workload and priorities

Schools were often very aware that structured conversations introduced an additional responsibility and time demand for teachers who were already managing heavy workloads and teaching responsibilities:

> *'[Teachers] are very receptive, but I do think it just comes back again to this absolute overload, the sheer huge workload people have now, and it is very hard because, you know, staff meeting time tends to get snaffled up for, you know, "we've got to do this, we've got to do this, we've got to do this"'* (champion, school 5).

In larger schools with several classes engaging in structured conversations, the distribution of target pupils was not always equal; as a result, some teachers had very few structured conversations to deliver whereas others had many:

> *'I think one of the issues we have found is how many children there are on the programme. I think that's been quite overwhelming and because we chose them from the year group as a whole some teachers have got far more children than other teachers […] so you've got some teachers who have got, like, ten children, others who have got two'* (champion, school 7).

Although it was seen as useful when the SLT facilitated these sessions by providing cover to relieve teachers from general classroom duties, teachers also wanted to avoid taking time away from their pupils:

> *'You can see it's really worthwhile, but the payoff is something else and the school have been really good here because they've said it can be happening during lesson time, but then it means that you're not with your children'* (teacher, school 5).

While many teachers felt that this additional task, though not always ideal, was still worthwhile, others felt that the conversations were not a major priority for them and so were a waste of their time: 'I just, honestly, felt like my hour was wasted when I could have been marking my books, to be quite honest' (new champion, school 8).

Perceptions of impact

It should be noted that several champions, coaches, and teachers reported a difficulty in assessing whether changes in practice were a direct result of AfA. They highlighted that the programme focused on developing existing areas of practice, rather than delivering something new: 'It's very difficult to isolate it' (champion, school 5); they also noted that AfA had been delivered alongside a number of wider efforts for improvement: 'I mean, nothing is in isolation is it? So it's really difficult to look at the impact of something when there's everything else going on in a school' (champion, school 3). Therefore, schools were unsure of the extent to which some impact would have happened regardless of AfA activity:

> *'I'm not sure if that's just him [a given AfA target pupil] working really hard and his ability anyway or if it has … it's sometimes difficult to know if that's [AfA] had a direct impact or if he would have done that anyway regardless of the structured conversation'* (teacher, school 3).

Leadership and overall school practice

Most of the direct and concrete changes that schools observed as a result of AfA took place at a system level, developing general school and leadership processes and practices. These included increased monitoring and evaluation of teaching, the use of the coaching model across leadership practices—including reflective questioning within teaching observations and appraisals, improved systems for SLT communication, and greater recording of intervention activity. Some schools reported developing their policies, including marking and attendance: 'Attendance was an issue across the cluster so [champion 2] rewrote our attendance policy and looked at how she analysed data using some of the things that [the coach] brought us from another school' (champion, school 3). A number of schools delivered additional training as a result of areas for improvement identified through AfA, which they felt in turn resulted in change:

> *'Me and [the coach] talked about, sort of, staff development and staff CPD. It certainly highlighted that perhaps staff weren't aware of key systems and the SEND code of practice […] the following week we put in a staff INSET around the SEND code of practice, so there was an immediate impact'* (SENCo, school 7).

Several champions and teachers felt that they saw little impact beyond these broad-level changes: 'I don't think it's had any impact on the attainment of the children at all but I do think in terms of contributing to our school self-evaluation then I think it's been helpful' (champion, school 4). However, a small number of schools implemented very little change at the school system level or felt that there were very specific and narrow changes rather than broad system improvement: 'I personally don't believe [that the school has taken anything new on]' (coach, school 8).

Individual teacher practice

Numerous champions felt that implementing school-level change, particularly increased monitoring and the use of a coaching model with teaching staff, resulted in changes to teacher practice (such as marking):

> *'We realise that for some of our members of staff it's not good enough just to say, "Here's our marking policy, this is what we expect." We've got to actually, physically show them how to do it, give them a chance to do it under our supervision, because there were inconsistencies in what was going on with the children's books in terms of marking and feedback and presentation and things. So that was actually quite useful'* (champion, school 4).

A number of school champions and other SLT and middle-leadership staff were believed to benefit from time spent with the coach assessing their own practices: '[The champion's] leadership skills are growing as well because she's getting more CPD really I think, so we've done the coaching […] just ways of managing staff at that leadership level' (coach, school 5). Some classroom teachers reflected that time they spent discussing an area of practice with the coach gave them a fresh perspective and changed their thinking: 'I did feel like she did offer some suggestions which made me think, "Ah, I never looked at it that way before." So, in that sense it was quite positive' (teacher, school 8). However, other schools observed that development was constrained to very specific changes to school-level processes, with individual teachers gaining little in terms of their professional development: 'I've learnt nothing' (champion, school 4).

Concrete changes to teacher practice often arose through structured conversations. It was frequently reported that engaging with parents in this way gave teachers greater insight into the particular circumstances and needs of target pupils and their families, which allowed them to tailor their teaching practice and pastoral support:

> *'We've had [structured conversations] with parents and they've helped to inform what we're doing in classroom and most particularly with the children with IEPs, because it's that personalised provision. So that gives you the chance to deviate far more than you would perhaps with other children'* (teacher, school 5).

Indeed, more broadly, conducting structured conversations was seen as facilitating teachers in developing their general parental engagement and communication skills, particularly for NQTs: 'I think now they're talking to parents on a proactive basis' (champion, school 2). However, some schools felt that there were limited direct changes to teaching practice besides structured conversations: 'There's been nothing that has changed in our practice' (headteacher, school 5).

Parent engagement

Schools commented that structured conversations were beneficial for the parents who engaged, in a number of different ways. Champions and teachers felt that parents gained a greater understanding of their children's learning and the different things happening in school: 'We have positive feedback from parents because they said they're getting a better understanding where their children are up to' (champion, school 4). In several cases, schools felt that the conversations led to a general increase in engagement and communication with these families:

> *'There are families who weren't engaging with school and with their child's learning who are now engaged. They're sharing concerns that they've got, barriers that they've got which the school didn't know about before which they're now able to address and provide a really sort of supportive nurturing relationship to them'* (coach, school 5).

It was highlighted that the conversations could provide parents with a space to talk about their difficulties: 'She had a lot of worries and issues with herself as well so sometimes it's nice for the parents just to offload and have someone to talk to definitely' (teacher, school 4), and that they took on board the guidance and suggestions, resulting in positive changes at home (for example, reading and homework support): 'Some of the parents were using strategies that they'd suggested' (coach, school 3). However, some teachers felt that parents were not always able to do this:

> *'I said there were certain things to be really handy for them to do regularly at home, things like practice reading regularly […] She basically just said, "Look, I've got six kids. I haven't got time to be supporting at home'* (champion, school 8).

There were also instances when schools highlighted that although these changes were positive, the time, logistical, and financial demands involved in delivering structured conversations may have outweighed the benefit: '[Teachers] did feel it was positive […] but certainly for the amount of time, energy, money, and everything else I cannot see that it is value for money' (champion, school 4). Indeed, there were cases where champions and teachers felt that these conversations offered little real benefit: 'I can't see any impact whatsoever' (champion, school 4).

Pupil impact

There were reports that changes to teaching approaches benefited pupils; for instance, in one school that adopted a focus on developing 'resilient learning', the champion reflected:

> *'There's a couple of children down in three I think […] you can just see in their books, you can just see from their attitudes, you can just see they do look and feel just a lot more resilient […] I think that's the big thing you're seeing some of those children do that a lot more confidently'* (champion, school 2).

However, as in this instance, these comments were frequently focused on individual pupils or small subgroups within the class rather than improvements for all pupils. Often this was focused on target pupils. For instance, several schools reported improvements in target pupils' learning attitudes and behaviours: 'I think what it's done is change learning behaviours and engage children in homework and actually that's had an impact in mathematics' (champion, school 7). Some champions and teachers felt that improved parent involvement in learning was beneficial for pupils' learning: 'I think sometimes what happens in schools is children come to school and they learn and actually it is at a distance from the parents, so I think it [structured conversations] brings learning much closer to home' (champion, school 6). However, several were sceptical as to whether these changes would have happened naturally over time anyway, regardless of AfA activity and structured conversations: 'It's sometimes difficult to know if that's had a direct impact or if he would have [improved in English attainment] anyway regardless of the structured conversation' (teacher, school 3). Indeed, a number of champions and teachers felt that AfA activity across the board had not led to any direct impact on pupils' learning: 'I don't think it's had any impact on the children at all and if the children do well it will be because of the work we're doing in school' (champion, school 4). There were a small number of instances where teachers and champions commented on changes to practice that impacted on pupils with SEND: '[Structured conversations] is the one big really big significant piece of work that's impacted the most in my opinion on the way that staff are working with children with additional needs' (champion, school 5). However, staff did not feel that changes in how they supported children with SEND during the course of the programme were attributable to AfA: 'I don't think there'd been any less focus on these children had we not been part of the scheme' (champion school 5).

As noted previously, schools reported very little activity regarding pupils' wider outcomes. As a result, there was limited discussion of impact in this area. Two schools reported a small increase in attendance as a result of increased monitoring or implementation of new strategies such as collecting pupils from home: 'I have actually seen an improvement for the summer term [for target pupils]' (champion, school 3). Several schools reported that some structured conversations focused on pupils' emotional wellbeing, though it was unclear whether this translated into positive improvement in outcomes. Beyond these examples, schools made no comment on whether AfA was believed to have influenced their pupils' wider outcomes.

Sustainability

Continuation of programme

None of the eight case study schools indicated that they were planning to continue with the AfA programme beyond the trial period, for varied reasons. Some observed that the trial period had been sufficient: 'I don't think we would be [continuing the programme] because I think we've probably learnt enough in the last two years, to be honest' (champion, school 2). Others felt that they had not seen sufficient impact and so they would not be continuing with the programme on that basis 'unless things change drastically' (headteacher, school 5). A small number of schools commented that their budget was limited and that AfA did not warrant priority within their resources: 'Our trouble at the moment is budget' (champion, school 8). Some felt this reflected the decision of the SLT, which may have differed from the opinion of other staff members (such as the school champion): 'With everything else, so of course he now looks at the AfA data and thinks, "Well this is not good." […] He's making a judgement at the moment that's not accurate' (champion, school 5).

Several champions and SLT members indicated that they would withdraw from activity now if they had not already invested time and money in it: 'If it were my decision, I'd pull the plug now' (assistant headteacher, school 8).

Continuation of specific practices

Schools did indicate that there were practices they would continue to implement: 'I'd just take those elements of it, you know, that we've already got […] from it' (champion, school 3). Several schools were interested in continuing to make use of a structured conversation approach, though perhaps not on the same scale or in exactly the same way:

> *'You wouldn't do it to all parents because not all children need it; [… we] probably wouldn't make them as long: we might make them a bit more frequent but, yeah, probably wouldn't make them as formal'* (champion, school 2).

Some schools were planning to roll out the structured conversations as a general practice across the whole school, rather than only within the target year groups, 'not necessarily in every year group but certainly parents and children who meet the criteria' (champion, school 1). Beyond structured conversations, schools' continuation of practices generally reflected particular changes to policy and practice that had been put in place through the course of the programme. One school wished to continue using a needs analysis and action plan approach informed by the programme: 'We quite liked the action plan when we did that' (champion, school 3).

Schools often reflected on the extent to which AfA practices had become embedded within their general practice, which was often highlighted as a critical factor in the programme: 'AfA is about embedding a different form of culture' (coach, school 8). Some schools put particular effort into giving the new practices and changes they were implementing the time and attention to become embedded rather than introducing something new and then immediately looking for another area to work on: 'You have to, I believe, give time to things and allow things to become well embedded […]; that practice [structured conversations] is really quite embedded now' (champion, school 5). Others reflected that the programme had become just an additional thing to manage, rather than becoming integrated into their general practice: 'I feel like it's an add-on, it's just there at the moment' (teacher, school 8). Indeed in some cases it was felt that withdrawing from the programme would give schools the space and time to embed what they had learned: 'Everybody's shell-shocked by the amount of things that we try to change so I think next year some of what we might do it embed what we've got' (champion, school 6).

Factors affecting sustainability

The champion was seen as an important component in how sustainable the programme was, given that they were typically the main person leading and monitoring activity. As noted previously, there were instances when champions were absent or the role shifted, and in these cases activity often stopped. However, it was highlighted that ideally this should extend beyond the champion over time, with activity becoming embedded across everybody's practice rather than being driven by one person:

> *'I've another school, an EEF [AfA trial] school actually, where the champion has changed but actually the programme is still going. So there it was down to how it is embedded by that champion before they go, because if they are the only person moving everything along then, yes, if the champion changes it [AfA] goes because it is common for a champion to change'* (coach, school 8).

Schools experienced difficulties when staff left or new staff came on board; in these instances, a coaching visit was typically used to deliver a second training session in order to facilitate them in understanding the programme and delivering structured conversations: '[The coach has] offered to come in and repeat some of the training with the new AfA teachers' (champion, school 7). There were some comments around the critical role of the coach, in that a school could leave the programme and continue with particular practices that they had learned, but could not continue to develop internally in the same way without the coach:

> *'It's been absolutely imperative, I think; having somebody external who comes in and just has a conversation where they're not heavily involved is just mind-blowing, mainly because she signposts so well but it's just really good to have those few minutes that you know are safeguarded in your diary so that you can sit and have that*

conversation, start unpicking things a little bit. So, I do think it's good and I would suspect that if we try and move forward without that role it would kind of dwindle off quite quickly' (champion, school 7).

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. This result has a very high security rating.
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. This result has a very high security rating.
3.	All children and children eligible for free school meals (FSM) 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 target children made three months less progress in maths, on average, compared to target children in control schools. FSM children in Achievement for All schools also made two months less progress in reading compared to FSM children in schools that did not receive the intervention.
4.	The evaluation found that the programme did not improve pupils' self-esteem, goals and aspirations, perceptions of how supportive their families were, or the attendance of target children. However, children in Achievement for All schools were more likely to report that there was an adult in their school who cared about and supported them.
5.	The implementation of Achievement for All was not optimal and varied across schools. However, there was no evidence to suggest that this contributed to the negative findings. Some teachers identified significant resource demands which made implementing Achievement for All challenging.

Interpretation

The current study is the second independent randomised trial of AfA, following that reported by Churches (2016). It combines an RCT on the basis of ITT involving 134 schools and 6,338 pupils in Y5, with an IPE strand combining surveys and in-depth case study work with eight schools purposively sampled to provide variation in overall attainment and added-value in attainment for the AfA target group. Given the methodological limitations of the earlier trial (most notably, lack of statistical power and very high rates of attrition, which in combination would have led to zero padlocks in EEF's trial security classification system), this can be considered the only robust evaluation of the contemporary version of the programme published to date.

In terms of impact on academic outcomes, our findings were consistent—if somewhat unexpected when compared to the findings from the initial AfA pilot (Humphrey and Squires, 2011a, 2011b). Across our ITT and (AfA target and FSM) subgroup analyses, business as usual was reliably found to be *superior* to AfA. Furthermore, this pattern of findings was almost completely insensitive to changes in our statistical models (specifically, MI of missing data and the inclusion of additional covariates). Indeed, the only discrepancy between our main models and the sensitivity analyses was that the marginal (p = 0.055), non-significant, negative impact on mathematics in the main ITT model was statistically significant when MI was applied or covariates were included (p = 0.027 in both cases; presumably because of slight gains in statistical power associated with these approaches, relative to the complete case model). Of note is that there are clear parallels between our findings and those of Churches (2016), who also found small but statistically significant, negative impacts of AfA on attainment at the ITT and FSM subgroup levels. Thus, notwithstanding the aforementioned limitations of the Churches (2016) trial, the evidence base for the contemporary version of the AfA programme now consists of two independent RCTs that have found it to be markedly inferior to business as usual.

The associated effect sizes in our analyses ranged from g = -0.12 for reading in the ITT model to g = -0.19 for mathematics in the AfA target subgroup model, equating to two to three months *less* progress among pupils in the intervention arm of the trial. Rather than interpret these using (completely arbitrary) conventional thresholds (for example, Cohen, 1992), it is perhaps more useful and meaningful to consider them in the context of the broader evidence base for universal interventions for school-aged children and young people (as recommended by Hill, Bloom, Black, and Lipsey, 2008). Our findings place AfA at the first percentile in the distribution of intervention effect sizes for academic outcomes (see Appendix D of Tanner-Smith, Durlak, and Marx, 2018). In other words, every school-based intervention for which academic outcomes were measured across the 74 meta-analyses, covering more than 1,100 controlled studies (involving nearly 500,000 children and young people) in Tanner-Smith et al.'s (2018) review produced more favourable effect sizes than those identified in the current study. The AfA target and FSM subgroup findings are particularly noteworthy as they demonstrate the failure of the programme to address policy-relevant performance gaps

(Hill et al., 2008). As noted in the introductory section of this report, both subgroups are considered vulnerable given the evidence that they are likely to experience disadvantageous outcomes during and following school (for example, leaving school early, lower lifetime earnings, and increased health and social care costs).

Clearly, then, AfA cannot be recommended for further scale-up on the basis of the evidence presented here. We consider various possible explanations for our pattern of findings. Stame's (2010) triumvirate of programme/theory failure, implementation failure, and research failure provide a helpful organising framework.

Theory failure

In a situation of programme/theory failure, implementation is as expected and associated evaluation activities are robust, but null (or in this case, negative) results are observed because the theory of change underpinning an intervention is inherently faulty. On this note, we found little to support the **AfA theory of change** (Appendix A) across our quantitative and qualitative implementation analyses. This theory of change is based on the premise that AfA is a school improvement programme that focuses on four areas for school development with core modules to support these. The intervention is intended to be flexible, in line with the findings of the national pilot (Humphrey and Squires, 2011b, 2011a). Similarly to the early phase of said pilot, some case study schools in the current trial found the flexibility in the programme design beneficial, however, others found the range of options overwhelming and the flexibility made it difficult to see where to start (Humphrey and Squires, 2010). Case study schools had different approaches to identifying the target group of pupils and the extent to which changes in practice were focused on those groups or focused on the wider school population (other than the structured conversation). As with the national pilot, coaches and champions reported the importance of all teachers understanding the nature of AfA and not seeing it as a 'bolt-on' intervention but rather one that required an attitudinal change towards inclusion.

Insufficient or faulty implementation activity

In situations of implementation failure, the programme theory is sound, and evaluation is robust, and hence null or negative results are attributed to insufficient or faulty implementation activity. Notwithstanding the methodological challenges associated with capturing and documenting the implementation of an inherently flexible programme such as AfA, the evidence gathered in our evaluation suggested that implementation was variable. For example, only around one in four (approximately 23%) of the AfA schools that completed the implementation survey reported having completed at least two structured conversations with parents/carers of all AfA target group pupils in a given school year. This is in stark contrast to the national pilot, in which this figure exceeded 80% (though of course as already noted, national pilot schools benefitted from generous funding that enabled them to release teachers to complete structured conversations). Fidelity to the structured conversation model was also somewhat lower, being 63% in the current trial compared to 84% in the national pilot (Humphrey and Squires, 2011a). Coaching support in the trial was implemented with frequency, though was still quite variable, with most AfA schools (71%) having had at least 20 coach visits by the end of the implementation period, others (23%) having had between six and 19, and a minority (6%) having had no visits due to them withdrawing from the programme before delivery began (relative to an overall target of around 24 visits in two years). It is also worth noting the unexpectedly low number of schools (n = 12 compared to AfA's pre-trial estimate of n = 33) that achieved the AfA QM status. However, we observed no substantive associations between implementation variability and intervention outcomes in either the path analyses in which we explored whether AfA implementation mediated the relationship between schools' socio-demographic context and pupils' academic outcomes, or in the multilevel regression analyses, in which we examined the extent to which implementation of different components of the AfA intervention predicted pupils' academic progress.[30] Thus, while implementation was by no means optimal (insofar as this could be ascertained given the flexible nature of the programme), it does not appear that higher levels of implementation were associated with improved outcomes.

The **coaching model** is central in the theory of change and involved professionals with a background in school leadership working with a school champion to reflect on issues identified in the school's needs analysis and development plan. In the AfA programme it is anticipated that 12 coaching visits are completed per year and most of the case study schools took up all of the coaching visits. The pattern of visits varied from the four per term identified in the AfA programme. In some cases, more visits were given in the first term to get schools started with the programme

[30] The sole exception to this being the significant association between activity pertaining to the leadership and inclusion strand and KS2 reading outcomes; however, this was a *negative* association.

reflecting the challenge of schools understanding a flexible programme like AfA that was identified in the national pilot (Humphrey and Squires, 2010). The role of coaches was to provide training and help with needs analysis, identification of vulnerable groups, strengthening of the approach to parental conversations, data tracking, and to direct teachers to online resources ('The Bubble') to support CPD. Schools generally found coaches helpful, but some commented on a misalignment of relevant experience and school needs or type. There seems to be a tension surrounding the relative roles of coaches and schools in developing an action plan and seeing it through the life of the programme. This is evident in terms of coach and school views not always aligning, schools not always resourcing their plan adequately, and shifting agendas on the part of schools or coaches. Although schools made use of the coaching visits to create school development plans, the school agenda and coach agenda did not always coincide. Coaches did not always consider the school plans to be making best use of their expertise. Termly reviews were often delayed due to a lack of time by school staff to analyse data. Some visits were cancelled due to staff absence or high workloads and schools did not rearrange visits with the coaches. Some schools reported a lack of follow-up on coaching visits that reduced the impact of the coaching on school development. There is some evidence that in some schools the action plan resulting from the initial needs analysis was not adhered to or adequately resourced. Schools seemed to be reacting to shifting external agendas (such as Ofsted inspections) that threw up new areas of need for development and a drift in focus away from AfA. This meant that some schools identified one area for development in their plan and then shifted to another area during the AfA intervention. A shift in focus for the wider outcomes from attendance to wellbeing may have also been introduced by the AfA coaches who encouraged schools to take up a new module that had been developed. For those schools, it could be argued that the action plan was not being used in a proactive way.

It would be expected that school improvement programmes would involve the **senior leadership team** (SLT) of the school and those with key responsibilities for target groups such as curriculum leaders and the SENCo. This seems to have been the case in the case study schools. In some schools there was a clear extension of the awareness of AfA to wider governance that included the SLT and the governing body. Champions were more effective when they were members of the SLT and AfA was prioritised as a school development area. Where members of staff had roles within an AfA school and roles within schools not involved in AfA (for example, in federation schools or academy trusts) then some of the staff reported extending ideas from the AfA school to the other school. In some schools, AfA champions felt that the level of engagement by the SLT with the coach was not as great as they would have liked, and it was left to the school champion to lead on implementation. Other teachers and teaching assistants in the school received initial training from the AfA coach but contact beyond this was limited.

Insufficient staff time resources required to implement AfA was raised in some of the case study schools and this resulted in school champions not always being able to meet with the AfA coach. In small schools, there was less flexibility as staff already had several competing roles and staffing issues had a greater negative impact on implementation. In larger schools there was more flexibility, but the target group of pupils was larger and there were tensions between working with 20% of a cohort, wanting to focus on a smaller number of pupils, or thinking about whole-school development and thinking about inclusion more generally.

The main feature of AfA that differentiates the programme from usual practice reported by teachers was the **structured conversations**, an aspect that provided an opportunity for more in-depth conversation with parents than would be achieved through parents' evenings. All of the schools set up structured conversations, and this had been an effective component identified in the national pilot (Humphrey and Squires, 2011a). When parents attended structured conversations, the school champions generally reported that these had gone well. In some cases, schools saw structured conversations as being beneficial to the extent that they could replace other processes such as individual education plan meetings and teacher consultations at parent evenings. Some of the ways that structured conversations were used allowed schools to gain insights into wider aspects of pupils' lives and provided parents with an opportunity to discuss problems that impacted on family life and school. Some schools were able to provide clear examples of how structured conversations had been beneficial in responding to individual needs of pupils. Coaches helped to improve fidelity of the parental conversations by providing further guidance to enhance the structure. However, the time needed for the structured conversations was not always made available by schools and in some cases staffing changes led to a lack of continuity. High workloads in some schools also reduced the ability of the school to complete parental conversations on a termly basis. The national pilot led to schools acting in a creative way to engage parents of children in the target group using additional resources provided (Humphrey and Squires, 2011a). This meant that many schools in the pilot were able to engage parents that are traditionally hard-to-reach. There were no additional resources for case study schools in the current trial and many found it difficult to engage hard-to-reach parents—in some cases, schools

were unable to get parents to attend the structured conversations. While we have evidence from the case study schools of champions experiencing frustration of not being able to engage some of these parents, we do not have any evidence of how schools problem-solved around this or provided alternative arrangements to meetings in school during the school day. Structured conversations placed high workloads on teachers that appears to have not been considered in the resourcing of AfA when schools agreed to take on the intervention.

One of the components from the national pilot that is difficult to replicate is how schools can network with other schools undertaking AfA. This was possible in the pilot because groups of schools within each of the ten participating LAs were geographically close and regional events were organised within the LA by the DfE or by National Strategies (Humphrey et al., 2013; Humphrey and Squires, 2011b). Under the current arrangements for AfA, schools opt into the intervention of their own choosing and are not recruited as a locality-based cohort. When there were two or more schools in the same locality, the AfA coaches were able to link the schools up directly and to set up local events and this allowed contact between key staff. For many schools this was not possible due to distance.

AfA has a loose theory of change that involves a flexible approach to school development supported by an external coach and this makes it difficult to identify an essential ingredient of the intervention. In many respects, the response of schools to issues identified as part of the AfA audit is similar to the response to any issues raised by external agencies. The focus of the intervention is on a target population identified by the school through a needs analysis and development of an action plan. There is a tension between viewing AfA as an intervention to support a target group and seeing AfA as a whole-school development to improve inclusion and outcomes for all learners through an attitudinal change. One component that appears unique to AfA is the structured conversation with the parents/carers of pupils in the target group; however, other similar approaches are possible such as Person-Centred Planning (Sanderson, 2015).

Research process

In situations of research failure, the programme theory is sound and implementation is as expected but problems with the design, data generation, and/or data analysis lead to an inability to detect a genuine intervention effect. Study limitations are noted in the section below with this in mind. Above and beyond these general methodological and analytical issues, it is noteworthy that the implementation period for the current trial was less than the two years recommended by AfA. Indeed, as noted earlier in this report, a planned addendum analysis of data pertaining to the cohort of pupils who were in Y4 at the beginning of the trial (who *will* have been exposed to two plus years of AfA, unlike their older counterparts) was set as part of the overall evaluation design for this reason. However, while insufficient intervention 'soak time' could potentially explain *null* results, it seems very unlikely that this would produce the *negative* effects observed here. A more feasible explanation, given the pattern of findings in the impact evaluation, is that schools in the control arm responded to not having been randomised to implement AfA by 'upping their game' in key areas of activity (for example, so-called 'compensatory rivalry'; Conrad and Conrad, 2005). However, we found only minimal evidence that this was the case—recall that we identified no significant changes over time by group (AfA versus business as usual) relating to the overall UPS score, or the 'leadership for inclusion', 'parent engagement', and 'wider outcomes and opportunities' subscales. The only exception was the 'teaching and learning' UPS subscale, where AfA schools reported a reduction in their CPD engagement over time, relative to an increase among control schools. Could this have contributed to the pattern of findings reported earlier? In our view, it is possible, but unlikely; after all, sensitivity analyses that incorporated UPS data produced the same effects as our main analyses. In other words, even when usual practice activities were taken into account, AfA still had a negative impact on our outcome variables.

Above and beyond issues pertaining to programme theory, implementation, and research process, it is also noteworthy that the key **target group(s) may have evolved**. Within the AfA programme there are arguably two approaches to defining target groups that reflect the complexity of school systems and the flexibility of the intervention. Firstly, the intervention is designed as a school improvement intervention, which would be expected to impact on the whole school population. This is achieved through four modules aimed at different groups in the school community. For example, 'leadership for inclusion' will aim to develop the SLT and governors; 'teaching and learning' will aim to improve how teachers deliver the curriculum; and, 'engaging with parents and carers' will aim to improve how teachers and parents work together. In the case study schools, there was some evidence of the target groups changing or not being able to engage: for 'leadership for inclusion', this included AfA champions being part of the SMT but then absent for extended periods and there was the situation where the headteacher was new to the post at the start of the trial. For 'teaching and learning', the initial focus was all teachers but then shifted to coaches working with curricular leads. Staffing

changes during the trial had a greater negative impact in smaller schools. Some schools reported that AfA was a useful way of helping staff adopt new roles while in other schools taking on new roles reduced the time available to commit to AfA. Reports from the case study schools suggest that skill levels increased amongst school staff, for example by improving monitoring and evaluation of teaching, developing marking and attendance policies and practice, developing reflective practice to gain fresh insights into teaching and learning, and improving conversations with parents.

Secondly, at the start of the AfA intervention, the coach worked with the school champion to identify a group of pupils who are vulnerable to underachievement. This group could be the lowest 20% of pupils based on assessment data or it could be another group that the school believes are vulnerable for other reasons (for example, FSM, travellers, migrant children, SEND, high mobility of children in military families, and so forth). In our analysis, the impact on this target group was based on the pupils identified by schools at the start of the programme, prior to randomisation. However, it is possible that schools changed the pupils that they wished to focus upon. For example, in some of the case study schools there was reluctance to focus on children with high mobility as they considered that they might start an approach with a child who then left the school.

Having considered the predominantly negative findings of this trial, in the interest of balance we should also note that **AfA led to improved school connectedness**. The aforementioned meta-analysis of the distribution of effect sizes for universal, school-based interventions did not focus on school connectedness specifically (Tanner-Smith, Durlak and Marx, 2018). However, taking perhaps the closest proxy, prosocial attitudes, $g = 0.15$, places AfA at the 50th percentile in the distribution of intervention effect sizes (that is, it produces larger effects than half of the interventions included in said meta-analysis). Converting this effect size to 'months' progress' to facilitate practical interpretation as per academic outcomes is clearly inappropriate. Instead, employing the U^3 metric (Durlak, 2009) indicates that AfA led to a six percentile improvement in pupils' perceptions of school connectedness. Given the developmental significance of this construct (for example, school connectedness is associated with higher academic achievement, general healthy development, and less mental health difficulties and risky behaviour; Panayiotou, Humphrey, and Hennessey, 2019), even this relatively modest improvement is clearly to be welcomed. However, we would of course caution against this as a rationale for the adoption of AfA given our other findings.

Strengths and limitations

This study has numerous strengths, increasing confidence in the security of our principal (impact) findings. A cluster-randomised design with appropriate analysis that took account of the hierarchical and clustered nature of the dataset was used. The trial was very large and well-powered, with an MDES of 0.14 at the point of randomisation. In terms of generalizability, the 134 trial schools spanned 78 of the 343 LAs across England. Attrition was 0% at the school-level and 8.28% 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. Notwithstanding the 'teaching and learning' CPD finding noted earlier, there was no substantive evidence of differential uptake of concurrent interventions. The use of a cluster-randomised design and the proprietary nature of the AfA programme minimised the possibility of contamination effects; moreover, there was no evidence of experimental effects (for example, Hawthorne or John Henry effects). Finally, 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 for obvious reasons it was not possible to achieve blinding in the administration of the tests).

Nonetheless, a number of limitations also need to be considered. First, given the non-manualised and inherently flexible nature of the AfA programme, capturing and documenting implementation was more challenging than is perhaps usual. This means that analyses that make use of quantitative IPE data should be considered exploratory as opposed to definitive. Second, on a related note, we were not able to apply complier average causal effect estimation (CACE) or related instrumental variable approaches to more robustly account for the role of implementation variability in intervention outcomes. CACE relies on a single, binary proxy for intervention compliance; this can be demanding even when working with a simple, manualised intervention, given the range of possible implementation dimensions (for example, dosage, fidelity, quality), but is nigh on impossible when evaluating a complex, flexible programme. Related to this, another limitation is the flexible and diffuse nature of AfA's theory of change, which made it difficult to identify

the ‘key features’ of the programme. Finally, AfA is proposed as a two-year intervention, yet the main (Y5) cohort analysed in this report were only exposed to one year and two terms of the programme. It is possible that acceleration of progress in academic outcomes such as reading and mathematics may only occur following two full years of exposure due to teachers gaining a better understanding of the programme over time. In support of this proposition, we note that in the national pilot, the proportion of structured conversations completed was higher in the second year of implementation (Humphrey and Squires, 2011a). Ultimately, the question of intervention exposure time will be addressed in the addendum to the report planned in 2020 in which the main analyses will be conducted with the data of the younger cohort (Y4) of pupils who were subjected to two full years of AfA.

Future research and publications

The findings reported here will also be summarised and disseminated via the Manchester Institute of Education (MIE) ‘Building Evidence into Education’ blog. We also plan to produce a series of additional outputs in academic journals (for example, Journal of Research in Educational Effectiveness). First, we will examine the relationship between resilience and academic success among primary school children. We will analyse attainment (that is, reading and mathematics) data of pupils in control schools to investigate possible mediating effects of resilience-related outcomes as measured by subscales of the SRS. Second, we will further investigate the negative effect of the intervention on academic achievement. In particular, drawing upon the rich qualitative IPE dataset, we aim to explore in more depth whether the hidden resource demands of AfA (for example, teacher time to complete structured conversations and engage with The Bubble) have created an aspirational model of implementation that is simply not feasible for schools.

References

Achievement for All 3As (2015) 'Impact Report 2014/15'.

AfA 3As (undated) 'Social Impact Assessment Schools Engagement Theory of Change', p. 10.

Akbulut-Yuksel, M. (2017) 'Do Legal School Leaving Rules Still Affect Schooling and Earnings?', *Social Science Research*. DOI: 10.1016/j.ssresearch.2016.06.013

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

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.

Borgna, C. and Struffolino, E. (2017) 'Pushed or Pulled? Girls and Boys Facing Early School Leaving Risk in Italy', *Social Science Research*. DOI: 10.1016/j.ssresearch.2016.06.021

Braun, V. and Clarke, V. (2006) 'Qualitative Research in Psychology Using Thematic Analysis in Psychology', *Qualitative Research in Psychology*, 3 (2), pp. 77–101. http://www.tandfonline.com/action/journalInformation?journalCode=uqrp20%5Cnhttp://www.tandfonline.com/action/journalInformation?journalCode=uqrp20.

Churches, R. (2016) *Closing the Gap: Test and Learn*. London.

Cohen, J. (1992) 'A Power Primer', *Psychological Bulletin*, 112 (1), pp. 155–159.

Conrad, K. M. and Conrad, K. J. (2005) 'Compensatory Rivalry', in *Encyclopedia of Statistics in Behavioral Science*, Chichester, UK: John Wiley. DOI: 10.1002/0470013192.bsa119

Department for Children Schools and Families (2008) *The Children's Plan One Year On: A Progress Report*, Nottingham: DCSF Publications.

Department for Children Schools and Families (2009a) *Achievement for All: Guidance for Schools*, Nottingham: DCSF Publications.

Department for Children Schools and Families (2009b) *Achievement for All: Local Authority Prospectus*, Nottingham: DCSF Publications.

Durlak, J. A. (2009) 'How to Select, Calculate, and Interpret Effect Sizes', *Journal of Pediatric Psychology*, 34 (9), pp. 917–928. DOI: 10.1093/jpepsy/jsp004

Durlak, J. A. and DuPre, E. P. (2008) 'Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation', *American Journal of Community Psychology*, 41 (3–4), pp. 327–350. DOI: 10.1007/s10464-008-9165-0

European Agency for Special Needs and Inclusive Education (2016) 'Early School Leaving and Learners with Disabilities and/or Special Educational Needs: A Review of the Research Evidence Focusing on Europe', edited by D. A. Dyson and G. Squires, Odense, Denmark.

European Agency for Special Needs and Inclusive Education (2017a) 'Early School Leaving and Learners with Disabilities and/or Special Educational Needs: To What Extent is Research Reflected in European Union Policies?', edited by G. Squires and D. A. Dyson, Odense, Denmark.

European Agency for Special Needs and Inclusive Education (2017b) 'Early School Leaving and Learners with Disabilities and/or Special Educational Needs: Final Summary Report', edited by G. Squires, Odense, Denmark.

Education Endowment Foundation (2013) 'Pre-Testing in EEF Evaluations', London: EEF. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/Pre-testing_paper.pdf

Farrell, A. D., Henry, D. B. and Bettencourt, A. (2013) 'Methodological Challenges Examining Subgroup Differences:

Examples from Universal School-based Youth Violence Prevention Trials', *Prevention Science*, 14 (2), pp. 121–133. DOI: 10.1007/s11121-011-0200-2

Fixsen, D. L. et al. (2005) 'National Implementation Research Network – Research Synthesis 2005'.

Hedges, L. V. (2007) 'Effect Sizes in Cluster-Randomized Designs', *Journal of Educational and Behavioral Statistics*. DOI: 10.3102/1076998606298043.

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.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D. et al. (2014) 'Better Reporting of Interventions: Template for Intervention Description and Replication (TIDieR) Checklist and Guide', *BMJ*, 348, p. g1687. DOI: 10.1136/BMJ.G1687

Hox, J. J. (2010) *Multilevel Analysis: Techniques and Applications*, New York: Routledge.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (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., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: A Synthesis of the Literature', p. 62. **https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf**

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*, *Education*. Nottingham.

Humphrey, N. and Squires, G. (2011b) *Achievement for All national Evaluation: Second Interim Report*. Nottingham.

Kreft, I. G. G. (1996) 'Are Multilevel Techniques Necessary? An Overview, Including Simulation Studies', pp. 1–26.

Lamb, B. (2009) *Lamb inquiry - special educational needs and parental confidence*. Nottingham.

Lendrum, A., Barlow, A. and Humphrey, N. (2015) 'Developing Positive School-Home Relationships Through Structured Conversations with Parents of Learners with Special Educational Needs and Disabilities (SEND)', *Journal of Research in Special Educational Needs*. DOI: 10.1111/1471-3802.12023

Lendrum, A. and Humphrey, N. (2012) 'The Importance of Studying the Implementation of Interventions in School Settings', *Oxford Review of Education*. DOI: 10.1080/03054985.2012.734800

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*, BioMed Central, 10 (1), p. 44. DOI: 10.1186/s13034-016-0132-5

Maas, C. J. M. and Hox, J. J. (2004) 'The Influence of Violations of Assumptions on Multilevel Parameter Estimates and Their Standard Errors', *Computational Statistics and Data Analysis*, 46 (3), pp. 427–440. DOI: 10.1016/J.CSDA.2003.08.006

OECD (2010) *Overcoming School Failure: Policies That Work*, OECD Publishing.

OECD (2012) *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*, OECD Publishing. DOI: 10.1787/9789264130852-en

Panayiotou, M., Humphrey, N. and Hennessey, A. (2019) 'Implementation Matters: Using Complier Average Causal Effect Estimation to Determine the Impact of the Promoting Alternative Thinking Strategies (PATHS) Curriculum on Children's Quality of Life', *Journal of Educational Psychology*. DOI: 10.1037/edu0000360

Price Waterhouse Cooper (2016) 'Achieving Schools: Social Impact Assessment Final Report 2016', London, UK.

Sanderson, H. (2015) *Progress in Preparing for Adulthood: Checking Your Progress in Delivering Personalised Support for Young People in Transition*, Stockport: HSA.

Squires, G., Humphrey, N., Barlow, A. and Wigelsworth, M. (2012) 'The Identification of Special Educational Needs

and the Month of Birth: Differential Effects of Category of Need and Level of Assessment', *European Journal of Special Needs Education*. DOI: 10.1080/08856257.2012.711961

Stame, N. (2010) 'What Doesn't Work? Three Failures, Many Answers', *Evaluation*. DOI: 10.1177/1356389010381914

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*, 19 (8), pp. 1091–1101. DOI: 10.1007/s11121-018-0942-1

Troncoso, P. (2020) 'Minimum Detectable Effect Size Calculator'. https://patricio-troncoso.shinyapps.io/mdesapp/

UN General Assembly (1989) *Convention on the Rights of a Child*, United Nations: Treaty Series, vol. 1577. DOI: 10.2307/4065371

UN General Assembly (2006) *Convention on the Rights of Persons with Disabilities and Optional Protocol*, United Nations: Sixty-first session, agenda item 67 (b). DOI: 10.1057/palgrave.development.1100310

UNESCO (1994) *The Salamanca Statement and Framework for Action on Special Needs Education*, Salamanca, Spain: United Nations Educational, Scientific and Cultural Organization.

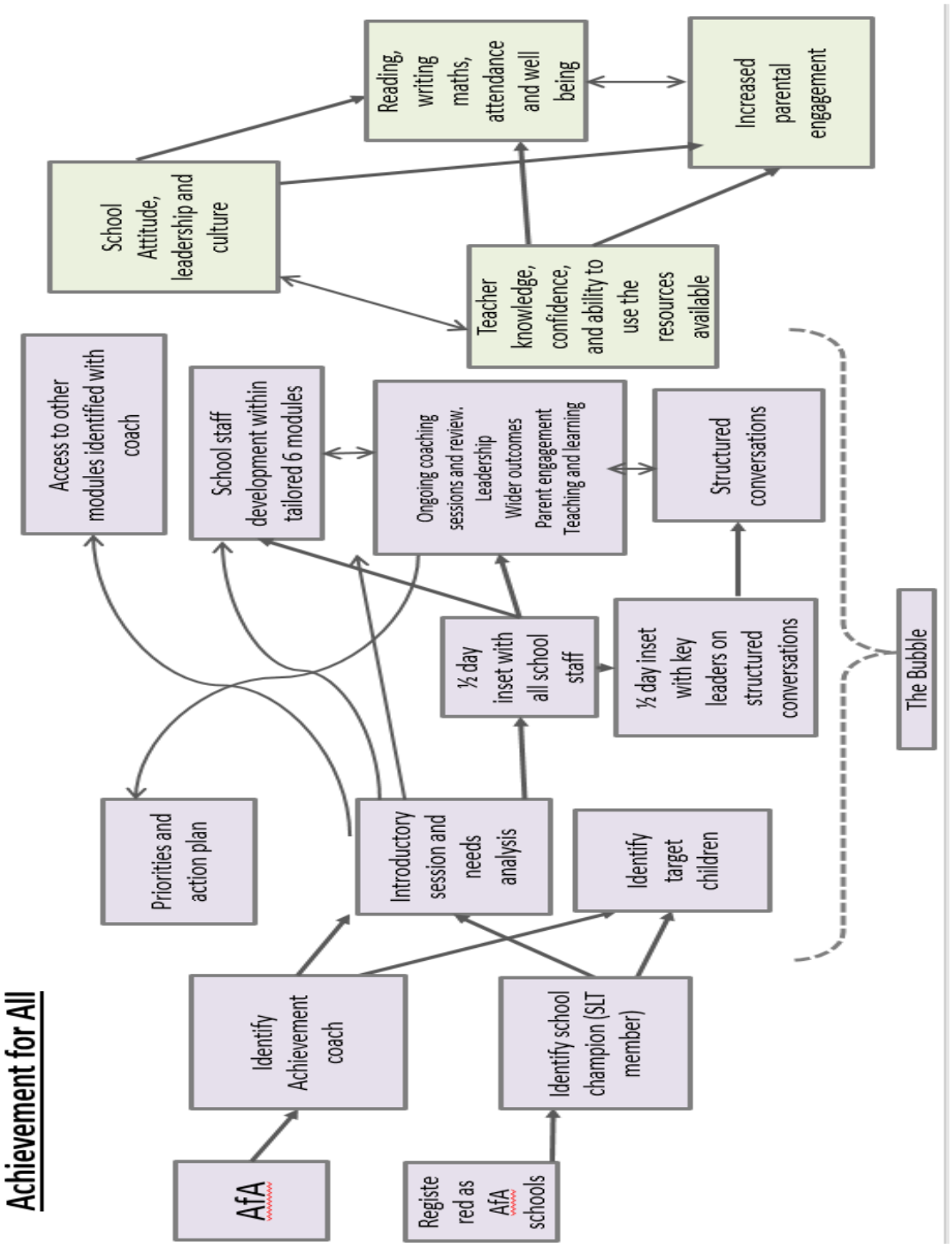
UNESCO (2000) *The Dakar Framework for Action. Education for All: Meeting our Collective Commitments*, *World Education Forum*, Paris, France: United Nations Educational, Scientific and Cultural Organization.

UNESCO (2015a) *Education 2030 Incheon Declaration: Towards Inclusive and Equitable, Quality Education and Lifelong Learning for All*, World Education Forum, United Nations Educational, Scientific and Cultural Organization.

UNESCO (2015b) *Relationship Between Sustainable Development Goal 4 and the Education 2030 Framework for Action*, United Nations Educational, Scientific and Cultural Organization.

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

Appendix A: Schematic of the AfA theory of change agreed at the start of the trial by AfA 3As, EEF and University of Manchester



Appendix B: Interview schedules used in case study schools

AfA champion

First Interview (Preparing and foundations) - Core open-ended questions

1. Can you briefly describe your experience in education?
2. Can you briefly describe your role and position in your school?
3. Can you describe your experience as an AfA school champion?
4. How do you describe the process of being assigned in this role?
5. Can you outline the main activities once you have taken up this role?
6. (Referring to the 'needs analysis') – what area of priorities have been identified your school? How do you identify the critical issues?
7. What actions or activities have been planned so far? Why? How do these address the priorities from the 'needs analysis'?
8. Why do you select certain activities? How does that relate to activities that your school is already involved with in relation to extra targeted support? Can you give a few examples?
9. What is the capacity for change in your school?
10. How do you ensure quality and effective delivery of the AfA program?
11. How many visits have you had from your AfA coach so far? What have you learned or implemented from these visits? If not, what are the concerns that have prevented implementation?
12. Are there any voluntary activities that you may be part of (as part of the AfA program)?
13. Can you describe your role in the school networks (if involved)?
14. If you are involved in networks, who does the responsibility of the network lie with?
15. If you are involved in networks, how do you share this responsibility with other champions? What are the expectations form you (and other champions)?
16. If you are involved in networks, what are the expectations from other school champions taking part in the network?
17. What are the difficulties and challenges generally?
18. What are the critical moments in this partnership (AfA coach and champion)? Why? What do you perceive to be the similarities and/or differences between the AfA program and other CPD opportunities the school might be involved with?

Subsequent Interviews (Term 2/3/4/5) – (review needs analysis document)

To begin with, have there been any changes in the AfA Champion/SLT roles generally and in relation to involvement with the AfA program?

Quality – how well different components of an intervention are delivered

1. Can you describe the activities you (or your lead teacher) have engaged with so far? How and by whom has this been delivered? What did you (or your lead teacher) learn from the activities? And how is your learning related to the delivery of the programme or your practice? Why did you choose those activities? Any difficulties or challenges?
2. Can you briefly describe the process of training and engaging with 'parental structured conversation'?
    a) Have you had a review with or without your AfA coach?
    b) How did you adapt the delivery in the previous term (different to what they were told by coach)?
    c) How would you change it in the future?
3. How effective has it been? Can you give a few examples of any outcomes due to the direct engagement with any AfA related activity or material? Did it work?
4. What are you (or your lead teacher) doing differently now as a result of the intervention? What's different? Can you give a few examples?

Dosage – how much of the intended intervention has been delivered and/or received

5. How often and in what form have you or your school received the intervention?
6. How often have you met with your coach and what was the focus?

Fidelity/adherence – the extent to which implementers (e.g. teachers) adhere to the intended treatment model

7. Can you outline any changes that have been implemented since the needs analysis? Why? Has your action plan changed?

Reach – the rate and scope of participation & **Responsiveness** – the degree to which participants engage with the intervention

8. Who else has benefitted from the AfA program so far? To what extent? In what ways? (external)
9. To what extent is the rest of the school staff involved in the program? In what ways? (internal)
10. How often have you engaged with 'The Bubble'? Anything else?
11. Has your AfA target group changed from the point of submitting data to UoM (July 2016)? How and why?
12. How often have you engaged with the coach? In what way and how did you find that experience? How has engaging with the coach helped your learning? In what way has that impacted your practice?

Programme differentiation – the extent to which intervention activities can be distinguished from other, existing practice

13. Are there any learning curves/moments that you have identified as part of the AfA program? What are these and why are they critical? How are these different from your usual CPD or training?

Adaptation – the nature and extent of changes made to the intervention

14. Are there any parts of the program that you felt the need to change (formally or informally) (e.g. parental structured conversation performa)? Why?

Refer to specific school issues identified in previous visits

Now thinking forward, what are the planned activities/plans for the next term/year?

AfA coaches

First Interview (Preparing and foundations) - Core open-ended questions

1. Can you briefly describe your experience in education?
2. Can you describe your experience as an AfA school coach?
3. How do you describe the process of being assigned to a school?
4. Can you outline the initial activities that you undertake, once you have been assigned to a school?
5. Why do you select certain activities? In your opinion, how does this relate to the activities that schools are already involved with in relation to extra targeted support? Can you give a few examples?
6. How do you ensure quality and effective delivery of the AfA program?
7. (Referring to the 'needs analysis') – what area of priorities have been identified in this school? How do you identify the critical issues?
8. What actions or activities have been planned so far? Why? How do these address the priorities from the 'needs analysis'?
9. How many times have you visited school x so far? What were these visits about?
10. Has there been any other sort of engagement with school x (online/email/etc.)?
11. What do you think is the capacity for change in the school that you are assigned to?
12. Are there any activities that you may be part of with as a school coach (outside of the core 12 visits)? For example, networks…
13. Can you describe your role in the school networks? How many networks are you involved with?
14. How do you share this responsibility with other coaches? What are the expectations form you (and other coaches)?
15. What are the expectations from other school champions taking part in the network?
16. What are the difficulties and challenges generally?
17. What are the critical moments in this partnership (between AfA coach and AfA champion)? Why?

Subsequent Interviews (Term 2/3/4/5) – (review needs analysis document)

To begin with, are you aware of any changes in the school specific roles (e.g. AfA Champion/SLT roles) generally and in relation to involvement with the AfA program?

Quality – how well different components of an intervention are delivered

1. Can you describe the activities your AfA school has engaged with so far? Why were those activities chosen? Any difficulties or challenges?
2. How did your school's parental structured conversation training and their delivery of the program go?
    a) Any challenges or successes?
    b) Are you aware of any changes that the school implemented?
3. Generally, how effective has the training been so far? Did you observe or are you aware of any outcomes due to the direct engagement with any AfA related activity or material?
4. How do you feel the school's practice has changed so far? Any specific examples.

Dosage – how much of the intended intervention has been delivered and/or received

5. How often and in what form have you delivered any direct intervention?
6. How regularly has your school engaged with program (either face to face or online)?

Fidelity/adherence – the extent to which implementers (e.g. teachers) adhere to the intended treatment model

7. Can you outline any changes that have been implemented since the needs analysis? Why? How were you informed (if at all)? Has your action plan changed?

Reach – the rate and scope of participation & **Responsiveness** – the degree to which participants engage with the intervention

8. To what extent is the rest of the school staff involved in the program? In what ways?
9. How often has your school engaged with 'The Bubble'? Anything else?
10. How did you find engaging with the school/AfA champion? How responsive was your school?

Programme differentiation – the extent to which intervention activities can be distinguished from other, existing practice.

11. Are there any learning curves/moments that you have identified as part of the AfA program in this school? What are these and why are they critical? In your opinion, how are these different from the school's usual CPD or training?

Adaptation – the nature and extent of changes made to the intervention

12. Are there any parts of the program that you felt the need to change (formally or informally)? Why?

Refer to specific school issues identified in previous visits.

Now thinking forward, what are the planned activities/plans for the next term/year?

Link teachers (if different from AfA champion) and other teachers

First Interview (measuring perceptions) Core open-ended questions

1) Can you briefly describe your role and position in your school?
2) Can you briefly describe your role within this AfA model?
3) How and why did you get involved?
4) What is the expectation from you in this model?
5) What are the expectations from others (school champion/coach)?
6) What do you perceive as the needs of the school? What is your focus? What are the activities you are involved with?
7) Attitudes and perceptions – what are your perceptions of the activities? What are your perceptions of 'The Bubble'?
8) How often do you anticipate engaging with 'The Bubble'?
9) What other activities so you anticipate of being involved with the AfA model but not directly related to it?
10) How is the agreed AfA program different from any other form of CPD you might be involved with?
11) What do you think are the intended outcomes of the AfA program?

Subsequent Interviews (Term 2/3/4/5) – (review needs analysis document)

Fidelity/adherence – the extent to which implementers (e.g. teachers) adhere to the intended treatment model

1. Can you outline any changes that have been implemented since the need analysis? Why?

Dosage – how much of the intended intervention has been delivered and/or received

2. How often and in what form have you or your school received the intervention?

Quality – how well different components of an intervention are delivered

3. Can you describe the activities you (or your lead teacher) have engaged with so far? How and by whom has this been delivered? What did you (or your lead teacher) learn from the activities?
4. What are you (or your lead teacher) doing differently now as a result of the intervention? What's different? Can you give a few examples?
5. How effective has it been? Can you give a few examples of any outcomes due to the direct engagement with any AfA related activity or material? Did it work?

Reach – the rate and scope of participation & **Responsiveness** – the degree to which participants engage with the intervention

6. Who else has benefitted from the AfA program so far? To what extent? In what ways?
7. How often have you engaged with 'The Bubble'? Anything else?
8. How often have you engaged with the coach? How has that helped your learning?

Programme differentiation – the extent to which intervention activities can be distinguished from other, existing practice

9. Are there any critical moments (unusual moments) that you have identified as part of the AfA program? What are these and why are they critical? How are these different from your usual CPD or training?

Adaptation – the nature and extent of changes made to the intervention

10. Are there any parts of the program that you felt the need to change (formally or informally)? Why?

Appendix C: Memorandum of Agreement

This Memorandum of Agreement outlines the key conditions for schools entering into partnership with Achievement for All in evaluation of the Achievement for All (AfA) Schools Programme. It outlines what schools that participate in the project will receive, and what they will be required to do in return. The aim is to have a completely transparent process so that all parties have a clear understanding of the project and shared expectations.

Section A: About Your School

We need some key details about your school – please complete the form below:

Name of school	
LAESTAB code	
School URN	
Address of school	
Postcode of school	
Telephone number of school	
Name of head teacher	
Email address of head teacher	

Section B: Your AfA Project Lead

This is the person who will co-ordinate the activity within the school and act as our first point of contact. This person needs to be a member of the Senior Leadership Team. Please provide details of the nominated lead person below:

Name of AfA Project Lead	
Email address of Project Lead	
Primary role within school	

Section C: Information about the UK trial

Aims of the evaluation

The aim of this project is to evaluate the impact of The Achievement for All Schools Programme, an approach that supports schools to improve outcomes in reading, writing and maths; improve behaviours and attendance; ensure inclusion and opportunity; and harness the full support of parents/carers. Ultimately, the programme not only improves the school experience and the life chances of children and young people, but also helps address many of the hurdles facing modern education including reduced capacity, ineffective personal development, and the need for strong leadership.

The evaluation is being conducted by the Manchester Institute of Education at The University of Manchester, and is funded by the Education Endowment Foundation (EEF).

The project

Following baseline data collection in the autumn term 2016, schools will be **randomly assigned** to a) Implement the AfA Schools programme or b) continue with their usual practice.
Schools randomly assigned to the AfA group will begin implementation in January 2017.

If assigned to the implementation group the schools nominated project lead will become the Achievement for All School Champion

Structure of the evaluation

A 2-year cluster-randomised evaluation will be used with randomisation at school level being undertaken by a statistician who is independent of the evaluator. Alongside this an implementation and process evaluation will also be undertaken. This means that all schools who decide to participate in the evaluation project agree that they can be **randomly assigned** to either (a) **implement the AfA Schools Programme**, or (b) **be a comparison school to continue their usual practice** over a **two-year period** (Jan 2017 – Dec 2018).

Random allocation is essential to the evaluation as it is the best way of determining what effect the AfA programme has on pupil outcomes. It is important that schools understand and consent to this process.

Section D: Key Conditions of Project Participation

In this section we outline the key conditions of project participation. Please read through them carefully.

All schools

Randomisation – all schools signing this document agree that they can be randomly allocated to either (a) implement the AfA Programme from January 2017 to December 2018 or (b) be a comparison school which continues their usual practice during this period. The randomisation procedure is scheduled to take place in early November 2016.

Focus – the evaluation cohort are pupils in Years 4 and 5 at the start of the 2016/17 school year.

Compliance with data collection requirements – all schools signing this document understand that they are committing to participation in an evaluation project with certain data collection requirements. These are:

(1) Provision of pupil information for the evaluation cohort (see above). As part of the process of signing up for the trial, schools agree to provide background data on all pupils in this cohort (including UPN, name, date of birth as well as personal characteristics and performance information). The data will also include the nomination of a 20% 'target group' that the school would like to focus on in the event that they are randomly assigned to implement the AfA Schools Programme. AfA will provide each school with a template file and guidance on selecting the 20% target group to facilitate this process.
(2) Distribution of information and consent sheets to parents of pupils in the evaluation cohort. The evaluation team will provide each school with copies of these brief documents.
(3) Completion of at least 90% of baseline (autumn 2016) and follow-up (summer 2018) surveys. We require pupils in the evaluation cohort to complete brief surveys that focus on resilience-related outcomes. The surveys will be administered through our secure online system and will take approximately 5 minutes to complete per pupil.
(4) Completion of school-level usual practice surveys at baseline (autumn 2016) and follow-up (summer 2018) by the project co-ordinator. The surveys will be administered through our secure online system and will take approximately 15 minutes to complete.

Schools randomly assigned to the AfA Schools Programme group:

For schools randomly allocated to implement the AfA Schools Programme, a commitment to implement the programme from January 2017 to December 2018 is required. AfA schools will be expected to sign a Service Level Agreement with Achievement for All and make a small contribution to the costs of implementing the programme. EEF is providing a 70% subsidy to the full programme delivery cost, this means the cost to schools is as follows:

- Dec 2016 £3,000
- Dec 2017 £2,000

Compliance with data collection requirements – in the event that they are randomly assigned to the AfA Schools Programme group, schools commit to the following additional data collection requirements:

(1) The AfA Champion in all AfA schools will be required to complete an implementation survey administered through our secure online system (summer 2018). This survey will take approximately 20 minutes to complete.
(2) A subsample of 7 schools in the AfA group will be recruited for longitudinal implementation case studies. In these schools, the evaluation team will collect additional data from a variety of sources (e.g. AfA coach, school AfA Champion, head teacher, class teachers, pupils, parents) using a variety of methods (e.g. observations, interviews, document analysis).

Schools randomly assigned to the comparison group:

Schools randomly allocated to the comparison group will continue practice as usual during from Jan 2017 to Dec 2018.

Section E: What Participating Schools Will Receive

This section outlines what each participating school will receive as part of the project.

All participating schools will receive:

Following each wave of pupil outcome surveys (autumn 2016 and summer 2018), the evaluation team will provide each school with a bespoke, aggregated feedback report that summarises the survey responses of

pupils in their school and compares them to those in the overall trial sample. This report may be a useful starting point for discussion within the school about possible needs of the pupil population and priority areas for development and improvement.

In addition, schools randomly allocated to the AfA Schools Programme group only will receive:

A comprehensive package of support to implement the Achievement for All Schools Programme. This will include:

- The allocation of an Achievement for All Achievement Coach who will undertake 12 half day visits to the school per year
- Working with colleagues from across the school community, the Achievement Coach and School Champion will develop an Action Plan that supports the implementation of the Schools Programme four elements (Leadership; Teaching and Learning; Wider Outcomes and Opportunities and Engaging with Parents and Carers) through high-impact and evidence-based core interventions, alongside and blended with a series of tailored school-driven activities.
- The delivery of a whole school Continuing Professional Development (CPD) session to introduce the wider school to the School Programme, its principles and approaches, Action Plan, whole staff engagement, data, and approaches to parent and carer engagement and roles.
- Selection of a target group of pupils in both years 4 and 5. These groups continue through into the second year of the programme, when schools will identify a further group of targeted pupils, based on the criteria in Y1.
  (N.B. detailed advice on the selection of the target group will be provided by Achievement for All)
- Use of data to improve pupil outcomes. A key aspect of the Schools Programme is the use of data to inform planning and improve pupil outcomes. We will therefore collect baseline data on pupils and teachers at the outset of the project and then data on progress and outcomes annually. Where practicable, we will collect quantitative data from progress assessments and qualitative data via pupil, parent and teacher focus groups.
- School wide access to The Bubble, Achievement for All's online community of practice delivering a huge range of resources to support programme implementation

In addition, schools randomly allocated to the comparison group only will receive:

A payment of £1000 will also be made to comparison schools for their participation in the evaluation project and the compliance with the data collection requirements outlined above. This payment will be staggered over the 2 years of the project as follows:

- £200 following confirmation of trial participation
- £200 at the end of the first year of the trial
- £200 at the midpoint of year 2 of the trial
- £400 at the conclusion of the trial and on completion of required data/surveys

Use of data

Pupils' test responses and any other pupil data will be treated with the strictest confidence. The responses will be collected online and/or on paper. The website that hosts the survey will be completely secure and password protected. All survey data will be stored on a secure, password protected computer to which only senior members of the research team have access. Named data will be matched with National Pupil Database data (e.g. on attendance and attainment) and shared with the EEF. No individual school or pupil will be identified in any report arising from the research.

Section F: Commitment to Participate

Application requirements

To complete their application for the AfA evaluation project schools need to:

- Send the completed Memorandum of Agreement to Achievement for All (details below) signed by the school's head teacher and Chair of Governors.
- Submit a completed evaluation cohort data file to Achievement for All by 31st July 2016
- Complete the baseline usual practice and pupil surveys, Sep-Oct 2016, prior to randomisation in the latter part of the autumn term 2016.

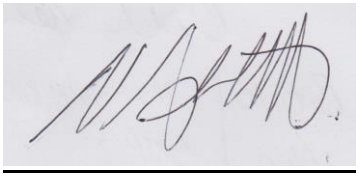
Commitment to participate

We confirm that we have read and understood all of the above and are happy to confirm our participation in trial of the Achievement for All Schools Programme as per the details specified, on behalf of

School: _____

_____	_____	_____
Head teacher (signature)	Print name	Date
_____	_____	_____
Chair of Board of Governors (sig.)	Print name	Date

On behalf of Achievement for All:

	Nick Aslett	25th April 2016
Project Lead, Achievement for All		Date

Please sign two copies, retaining one and returning the second copy to:

Achievement for All
St Anne's House,
Oxford Square,
Newbury,
Berks,
RG14 1JQ

Alternatively, please scan and email your completed form to: takepart@afaeducation.org

Appendix D: Information sheet for parents/carers

ACHIEVEMENT FOR ALL – Randomised Controlled Trial
INFORMATION SHEET FOR PARENTS

Your child's school has agreed to participate in a research project being run by the University of Manchester. We have been commissioned by the Educational Endowment Foundation to evaluate a programme called Achievement for All. The aim of Achievement for All is to improve children's learning and experience of school.

We are writing to you because your child's school is involved in the project. We will be making use of information that is routinely collected by the school about progress in English, Maths and attendance. We will ask for this information from the National Pupil Database. To see whether Achievement for All makes a difference or not, we will randomly allocate half of the schools to receive the programme and half of the schools will carry on as normal.

We will also ask pupils who are in Year 4 and Year 5 in the school year 2016/17 to complete a short questionnaire in Sep/Oct 2016 and again in May/Jun 2018 to see if Achievement for All improves their resilience.

Please take time to read the following information carefully and decide whether or not you would like your child to take part.

If you would like any more information or have any questions about the research project, please telephone Dr. Sophina Choudry on 0161 275 3534 or email her at: sophina.choudry@manchester.ac.uk.

Who will conduct the research?

The research will be conducted by Dr. Garry Squires and other staff in the School of Environment, Education and Development, University of Manchester, Oxford Road, Manchester M13 9PL.

Title of the research

Achievement for All – Randomised Controlled Trial

What is the aim of the research?

Our main aim is to find out what impact Achievement for All has on outcomes for children.

Where will the research be conducted?

In primary schools in England.

What is the duration of the research?

The project runs from September 2016 until August 2020.

Why have I been chosen?

We are writing to you because your child's school is taking part in the Achievement for All research project.

Schools normally review how they are approaching teaching and learning and then plan to make changes. Achievement for All is a scheme in which the schools can receive support in this planning from a coach. The Achievement for All coach will visit the school 12 times in each year of the project to work with teachers. The programme will focus on four key areas: School Leadership; Teaching and Learning; Wider Outcomes and

Opportunities; and, Engaging parents and carers. As part of the schools planning, there will be opportunities identified for teacher training and development. As well as thinking about all of the pupils in the school, teachers will be asked to focus on how a target group of Year 4 and Year 5 pupils that teachers think need more support.

In this research project we want to know if Achievement for All makes a difference to how well children do in English and Maths. We also want to know if it improves attendance and improves how children feel about themselves.

Schools will be randomly chosen to (a) implement the Achievement for All Schools programme over a two-year period (Achievement for All schools), or (b) continue as normal (comparison schools). This means that your child may be in an Achievement for All school or may be in a school that is continuing as normal.

We will be collecting data in both Achievement for All and comparison schools. After two years, all schools will be free to decide whether they wish to start/continue using the Achievement for All.

Your child is in one of the target year groups (Years 4 and Year 5 in the school year 2016/17) that we want to find out about.

What would my child be asked to do?

- Complete a short survey of 13 items that looks at how they feel about themselves, the involvement of their family in schooling, engagement with school, and their goals and aspirations.
- An example item is “I can do most things if I try” and the response is on a 5-point scale (where 1 = Never and 5 = Always).
- The survey will be completed twice: once in September/October 2016 and again in May/June 2018

We are giving you the opportunity to opt your child out of the study. If you do not opt out then you will be considered to be giving permission for your child to complete the survey. If you do not opt out but then change your mind later, we can remove your child’s data from our records.

There will be no direct contact between any of our research team and your child for this part of the research project.

In consenting to your child’s participation, you are also giving permission that for the purpose of the study, information provided will be linked with the National Pupil Database (held by the Department for Education), other official records, and shared with the Department for Education, Education Endowment Foundation (EEF), EEF’s data contractor FFT Education, and in an anonymised form to the UK Data Archive.

What happens to the data collected?

The data will be analysed by our research team at the University of Manchester. We will write a report based on our analyses for the Education Endowment Fund. It is also likely that we will write articles for academic journals based on what we find out in the project. Finally, it is possible that we will write a book about the research. Your child’s name will not be used in any of the reports that we write.

How is confidentiality maintained?

All data provided will be treated as confidential and will be completely anonymous. Identifying information (e.g. your child’s name) will only be used in order to match responses to data from the National Pupils Database on English, Maths and attendance. After this matching process is complete all identifying information will be destroyed.

The website that houses the survey will be completely secure and password protected. All survey data will be stored on a secure, password protected drive to which only senior members of the research team have access.

What happens if I do not want my child to take part or I change my mind?

It is up to you if you want your child to take part.

If you decide that your child can take part you do not need to do anything.

If you decide **not to allow your child to take part,** then you need to either complete the opt-out consent form enclosed and return it to our research team at the address above or contact Dr. Sophina Choudry by telephone or email (details below) by **7th October 2016.**

If you decide your child can take part and then change your mind, you are free to withdraw them from the study without needing to give a reason by contacting Dr. Sophina Choudry by telephone or email. We will send annual reminders about the study, but you can opt your child out at any time up until the end of the study, in summer 2019. If you do opt out, please rest assured that we will destroy any data collected about your child as part of the study.

Will I be paid for participating in the research?

We are not able to offer any payment or incentive for participating in this study.

Criminal Records Check

Every member of our research team has undergone a Disclosure and Barring Service check at the Enhanced Disclosure level.

Contact for further information

Dr Sophina Choudry
Manchester Institute of Education
University of Manchester
Oxford Road
Manchester
M13 9PL
Tel: 0161 275 3534
Email: sophina.choudry@manchester.ac.uk

Also, please see our website for further details: www.afatrial.info

What if something goes wrong?

If completing the survey makes you worry about your child's wellbeing then you should contact his/her school in the first instance and ask to speak to the Achievement for All project link teacher.

If you have further concerns then contact Dr. Sophina Choudry (details above).

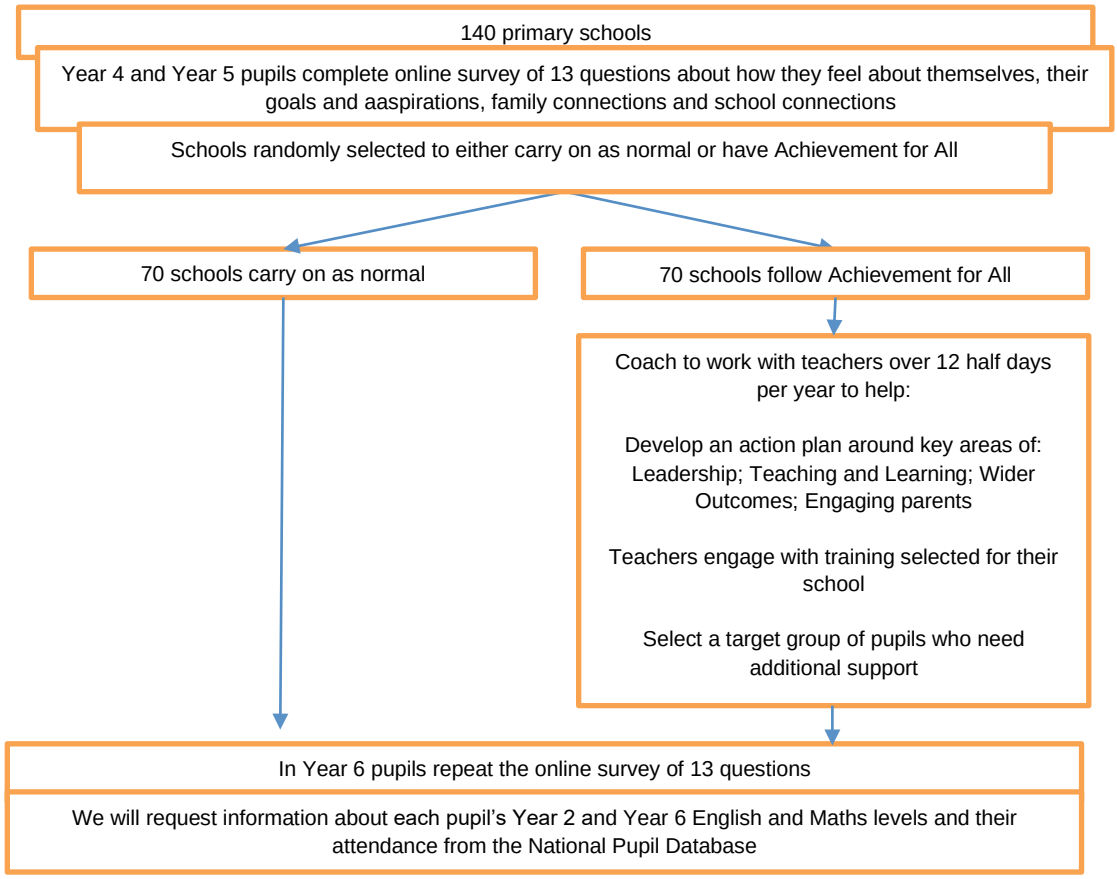
What if I want to complain?

If you wish to complain, you should contact Dr. Sophina Choudry in the first instance (contact details above).

If you remain dissatisfied, or if the research team is unable to address the issues you raise you should contact the Head of School, Prof Tim Allott (School of Environment, Education and Development), at Tim.Allott@manchester.ac.uk or on 0161 275 3662.

If there are any issues regarding this research that you would prefer not to discuss with members of the research team or Head of School, please contact the Research Practice and Governance Co-ordinator by either writing to 'The Research Practice and Governance Co-ordinator, Research Office, Christie Building, The University of Manchester, Oxford Road, Manchester M13 9PL', by emailing: research.complaints@manchester.ac.uk, or by telephoning 0161 275 7583 or 275 8093.

Achievement for All RCT summary for parents


140 primary schools
Year 4 and Year 5 pupils complete online survey of 13 questions about how they feel about themselves, their goals and aaspirations, family connections and school connections
Schools randomly selected to either carry on as normal or have Achievement for All
70 schools carry on as normal
70 schools follow Achievement for All
Coach to work with teachers over 12 half days per year to help:
Develop an action plan around key areas of: Leadership; Teaching and Learning; Wider Outcomes; Engaging parents
Teachers engage with training selected for their school
Select a target group of pupils who need additional support
In Year 6 pupils repeat the online survey of 13 questions
We will request information about each pupil's Year 2 and Year 6 English and Maths levels and their attendance from the National Pupil Database

ACHIEVEMENT FOR ALL – Randomised Controlled Trial
PARENTAL CONSENT FORM

An information sheet is attached to this form. Please read it carefully before making a decision about taking part in the study.

If you are willing to allow your child to take part then you do not need to do anything at the moment. In consenting, you are also giving your permission for your child to complete the 13 questions in the online survey.

If you decide not to take part, then you need to complete the opt-out consent form below and return it to:
FREEPOST RLYU-KAAB-AXRC
Dr. Sophina Choudry
Manchester Institute of Education
University of Manchester
Oxford road, Manchester
M13 9PL

Alternatively, Dr. Choudry can be contacted by telephone on 0161 275 3534 or email at sophina.choudry@manchester.ac.uk . If you do not wish to participate please let us know **by 7th October**.

Finally, please also remember that if you do decide to take part, you are free to change your mind at any point in the study up to the end of the study in July 2019. Just let us know and we will destroy any data generated in relation to your child.

I **do not** wish my child to participate in the Achievement for All research conducted by the University of Manchester.

Name of child	
Sex of child	
Year group	
Name of school	
Local Authority (if known)	
Signature (Parent/Guardian)	
Date	

Return this slip to:
FREEPOST RLYU-KAAB-AXRC
Dr. Sophina Choudry
Manchester Institute of Education
The University of Manchester
Ellen Wilkinson Building
Oxford Road
Manchester
M13 9PL

Appendix E: General Data Protection Regulation notice

General Data Protection Regulation Notice

Thank you for agreeing to your child's participation in the Achievement for All (AfA) effectiveness trial which started in 2016. Recently there have been changes to data protection law and this leaflet is to tell you about those changes.

This is to outline your and your child's rights with respect to processing of data. These rights are as set out in the General Data Protection Regulation (GDPR), which supersedes the Data Protection Act from May 2018.

The project will involve collecting data on pupils' attainment and background details, as well as their resilience. Pupils and schools will be asked to complete short surveys at two time periods: Sep/Oct 2016 and the current surveys running from April - May 2018. The surveys will be conducted in schools through a secure online platform. The responses will be accessed by the University of Manchester. For the purpose of research, the responses will be linked with information about the pupil from the National Pupil Database and shared with the Department for Education, the Education Endowment Foundation (EEF), FFT Education (EEF's data processor for the archive) and, in an anonymised form, with other research teams and potentially the UK Data Archive. Further matching to NPD data may take place during subsequent research. Your child's data will be treated with the strictest confidence using pseudo anonymised information in line with General Data Protection Regulation (GDPR).

The legal basis

The legal basis for processing these data for the research project is public interest (Article 6(1)(e) and Article 9(2)(j) of the General Data Protection Regulation). This means that personal data can be processed where necessary for the performance of a task carried out in the public interest. In this case it is to carry out research and inform future educational provision.

Your rights

The General Data Protection Regulation is designed to protect and support the following personal data rights for everyone in the UK:

- **The right to be informed**:
  - Your child's data is being processed by the project research team.
- **The right of access**:
  - You have already been provided with the Participant Information Sheet which explains what data we are collecting and what we are doing with it. The data will be used to produce a research report. The information sheet is also available on the website **www.afatrial.info/documents**.
- **The right to correct data:**
  - The right to correct incorrect records. All personal information has been collected from the National Pupil Database held by the Department for Education. This information is constructed from records held in school and if you want to change this data then you should contact your child's school. We will be analysing anonymised data.
- **The right to be forgotten:**
  - The right to request that data is removed/deleted. If you want to do this then please contact the research team.
- **The right to restrict processing:**

    - The right to request that data be held but not processed unless necessary. We will only process the data to answer our research questions. This is on the project information sheet, which is available on the website www.afatrial.info/documents.
- **The right to data portability**:
    - The right to a copy of your data in a useable format. Once the data has been anonymised we are unable to retrieve information for a specific child.
- **The right to object**:
    - You may object to your data being processed. You are free to withdraw your child from the research project up to the point at which the analysis starts.

Holding the data

During the evaluation, the organisation in control of personal data collected for this research is the University of Manchester. After the research has completed, Education Endowment Foundation (EEF) is the data controller for the Fischer Family Trust (FFT) Education archive. The University of Manchester is also responsible for collecting and processing the data from this project.

We will not be transferring any identifiable information outside the EU and will be taking appropriate measures to ensure it remains secure at all times.

We will keep the pseudo anonymised information, where individuals won't be readily identifiable, for a 4-year period while the research project is active. After that we will change it to make individuals in the data set completely unidentifiable. This anonymous information may then be used for research for another 5 years. After this, the information and data will be securely destroyed by the University of Manchester. EEF has their own data retention policy, which will be available shortly.

Please note that the consent obtained in 2016 for these surveys relate to involvement in the research but these are not the legal basis for data processing. As described above, the legal basis for data processing is public interest. Your data rights with regard to data processing have been set out in this notice and will be respected. For further information, please see https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/ and https://ico.org.uk/for-the-public/is-my-information-being-handled-correctly/.

If you have any concerns or questions about our research, the data processing, and/or your involvement in the project please contact:

Dr Sophina Choudry
Research Associate
0161 275 3534
Sophina.Choudry@manchester.ac.uk
Dr Garry Squires
Lead investigator
0161 275 3546
Garry.squires@manchester.ac.uk

Appendix F: Student Resilience Survey (SRS) for the AfA trial

	Never 1	Rarely 2	Sometimes 3	Often 4	Always 5
I can work out my problems					
I can do most things if I try					
There are many things that I do well					
I have goals and plan for the future					
I think I will be successful when I grow up					
There is an adult at home who is interested in my schoolwork					
There is an adult at home who believes I will be a success					
There is an adult at home who wants me to do my best					
There is an adult at home who listens to me when I have something to say					
There is an adult who really cares about me at school					
There is an adult at school who tells me when I have done well					
There is an adult at school who listens to me when I have something to say					
There is an adult at school who believes I will be a success					

Appendix G: The AfA Teacher Implementation Survey

Achievement for All Trial

AfA Teacher Implementation Survey 2017-18

This survey is designed to explore your implementation of the Achievement for All (AfA) Programme. We are interested in the feasibility of implementing AfA in English schools and the factors that may affect this, and not judging individual performance. The information you provide will be treated as anonymous and confidential. Please answer as honestly as possible. The survey should 10-15 minutes to complete.

Part 1 – ABOUT YOU AND YOUR ROLE(S) IN SCHOOL

1. Please indicate which of the following roles you undertake in your school (tick any that apply):

   - ❑ Head teacher
   - ❑ Deputy or assistant head teacher
   - ❑ Special educational needs co-ordinator/inclusion manager
   - ❑ Head of year or key stage
   - ❑ Class or subject teacher
   - ❑ AfA School Champion
   - ❑ Other (please specify) _____

2. Are you part of your school's leadership/senior management team? ☐Yes ☐ No

3. Has your AfA champion been in this role from the beginning of the programme? ☐ Yes ☐ No

Part 2 – IMPLEMENTING AfA IN YOUR SCHOOL

SECTION A – ABOUT IMPLEMENTATION

1. This question will help us understand *when* AfA has been implemented in your school. Tick as many as apply.

☐ Summer Term 2017
☐ Autumn Term 2017
☐ Spring Term 2018
☐ Summer Term 2018

2. If you stopped implementing during one or more terms (as indicated in the previous question), please explain why. [Question 2 to show only, if one or more is unchecked in previous question]

SECTION B – ABOUT LEADERSHIP FOR INCLUSION IN YOUR SCHOOL

1. Which of the following five 'leadership for inclusion' elements have your school chosen to focus upon?

   - ❑ Strategic planning
   - ❑ Governance
   - ❑ Staff performance and development
   - ❑ Community relationships
   - ❑ Pupil progress
   - ❑ None of the above [exclusive]

[SURVEY TO ONLY SHOW ASSOCIATED TABLES WITH SELECTED OPTIONS ABOVE]

2. Please rate the level of progress made by your school in relation to the following activities using AfA funding/resources:

	Strategic planning	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*How effectively do strategic teams currently set targets for improvement?*				
2	*How effectively do School Development Plans priorities focus on improvements in outcomes for the vulnerable pupils?*				
3	*How well is the School Champion developing leadership capacity e.g. sharing and applying coaching approaches with staff?*				
4	*How effectively does a designated member of the teaching staff co-ordinate and integrate multi-agency work and key working?*				
5	*How effectively are provision maps used across the school to ensure the strategic management of provision?*				

	Governance	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*Do governors receive regular updates on the progress of vulnerable groups and target pupils and the impact of current policies?*				
2	*How effectively has leadership of Achievement for All been distributed between staff, with clear lines of accountability?*				

	Staff performance and development	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*How regularly does Senior Leadership Team use observations and learning walks to support school improvement?*				
2	*How well does the Continuing Professional Development plan for teachers and support staff include training to support the Achievement for All Schools Programme?*				
3	*How effectively do staff performance targets incorporate expectations of consolidation and accelerated progress for the target pupils?*				
4	*How effectively are school leaders coaching and building leadership capacity?*				

	Community relationships	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*How effective is the communication between colleagues and other stakeholders in creating opportunities and improved outcomes for pupils?*				

	Pupil Progress	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*How effectively does the Senior Leadership Team secure teaching and learning that is consistently good or better?*				
2	*How effective are school monitoring and evaluation systems in driving improvements in teaching and learning?*				
3	*Are target group pupils making at least expected progress, and is the gap closing with their peers?*				
4	*To what extent do staff focus on the progress of the lowest achieving pupils?*				
5	*How effectively is progress data used by senior leaders and teachers to raise aspirations and drive progress?*				
6	*How robust are the school's internal and external moderation processes in ensuring secure and validated teacher assessments?*				

3. The next question is related to your staff's engagement with the core area Leadership for Inclusion of the online platform 'The Bubble'.

   Please answer by clicking on the number that best shows to what extent you have engaged with the specified Bubble module. If you have not engaged, pick a number close to 0. If you have fully engaged, pick a number close to 10. Work quickly, but carefully. There are no right or wrong answers.

		Not engaged at all										**Fully engaged**
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1.	*Please rate the extent to which your staff have engaged with the Leadership for Inclusion core area of the online platform 'The Bubble'. Example modules are coaching for inclusive leadership, leadership for inclusion, and provision to close the gap.*											
2.	*Please rate the impact of the Leadership for Inclusion core area of the online platform 'The Bubble' has had in your school. Example modules are coaching for inclusive leadership, leadership for inclusion, and provision to close the gap.*											

SECTION C – ABOUT TEACHING AND LEARNING IN YOUR SCHOOL

1. Which of the following four teaching and learning areas have your school chosen to focus upon in AfA?

   - ❑ Quality of teaching and learning
   - ❑ Effective use of interventions
   - ❑ Use of additional staff
   - ❑ Pupil voice
   - ❑ None of the above (exclusive)

   [SURVEY TO ONLY SHOW ASSOCIATED TABLES WITH SELECTED OPTIONS ABOVE]

2. Please rate the level of progress made by your school in relation to the following activities using AfA funding/resources:

	Quality of teaching and learning	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*Does every teacher show full commitment to improving the learning and achievement of every pupil in their class?*				
2	*What strategies are in place to support teachers in moving from good to outstanding?*				
3	*How well-established are coaching principles and practices?*				
4	*How regularly and effectively are interventions for vulnerable pupils monitored?*				
5	*How effectively is teacher feedback to pupils established and monitored across the school?*				
6	*Are all teachers able to identify Speech, Language and Communication Needs and aware of effective interventions?*				

	Effective use of interventions	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*Are teaching and learning strategies regularly informed by pupils' individual targets and are class teachers responsible for the selection, implementation/oversight and monitoring of interventions?*				
2	*How well are pupils' responses to personalised interventions monitored?*				
3	*Do teachers use evidence based interventions to address individual needs? Are effective interventions embedded into Quality First Teaching?*				

	Use of additional staff	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*How robust is the evidence that additional adults are having a positive impact on pupil progress and well-being?*				

	Pupil Voice	Fully embedded	In place	Partly in place	Not in place
1	*How well are the school's channels of expression working for its vulnerable pupils?*				

3. The next question is related to your staff's engagement with the core area Teaching and Learning of the online platform 'The Bubble'.

   Please answer by clicking on the number that best shows to what extent you have engaged with the specified Bubble module. If you have not engaged, pick a number close to 0. If you have fully engaged, pick a number close to 10. Work quickly, but carefully. There are no right or wrong answers.

		Not engaged at all										Fully engaged
		0	1	2	3	4	5	6	7	8	9	10
1.	*Please rate the extent to which your staff have engaged with the Teaching and Learning core area of the online platform 'The Bubble'. Example modules are provision to close the gap, max. impact of Teaching and Learning assistants, and using effective feedback.*											
2.	*Please rate the impact the Teaching and Learning core area of the online platform 'The Bubble' has had in your school. Example modules are provision to close the gap, max. impact of Teaching and Learning assistants, and using effective feedback.*											

SECTION D – ABOUT PARENT AND CARER ENGAGEMENT IN YOUR SCHOOL

1. How would you describe the school's relationship with parents of AfA target children *at present?* In terms of this relationship, we are particularly interested in the level of engagement and confidence among parents.

   - ❑ Excellent relationships with most or all parents
   - ❑ Good relationships with most or all parents
   - ❑ Poor relationships with most or all parents
   - ❑ Very poor relationships with most or all parents

2. What proportion of parents of children identified as AfA target students in the current Year 5 and 6 groups have taken part in structured conversations this school year:

[KEY SURVEY TO SET UP SO THAT THE TOTAL ADDS UP TO 100%]

No. of structured conversations	**Proportion %**
3 or more structured conversations this school year?	
2 structured conversations this school year?	
1 structured conversations this school year?	
No structured conversations this school year?	

3. The following questions are designed to assess the progress made by your school in engaging with parents and carers. Please read each statement and then indicate how much it applies to the activities relating to parent and carers' engagement that have taken place in your school:

	Inclusion	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1.	*Are a range of strategies in place to engage previously 'hard to reach' Parents and Carers?*				
2.	*How effectively have approaches to welcoming all families been established across the school?*				
3.	*Are Parents and Carers effectively and frequently updated on their child's progress with indication on how they can effectively support their child(ren) at home?*				

	Structured Conversations	**Always**	**Often**	**Rarely**	**Never**
1.	*School support key teachers develop the skills for Structured Conversations with Parents and Carers*				
2.	*Structured Conversation meetings are effective.*				
3.	*Structured Conversation meetings are frequent and well-attended.*				
4.	*Parents and Carers and staff meetings are collaborative in terms of agreeing targets, strategies, intervention and support.*				
5.	*Pupils' views are considered in meetings with Parents and Carers.*				

6.	*Pupils are involved in at least part of these meetings.*				
7.	*Systems to support staff, where appropriate, to signpost resources and training to Parents and Carers and young people are in place and accessed frequently.*				

4. The next question is related to your staff's engagement with the core area Parent and Carer Engagement of the online platform 'The Bubble'.

   Please answer by clicking on the number that best shows to what extent you have engaged with the specified Bubble module. If you have not engaged, pick a number close to 0. If you have fully engaged, pick a number close to 10. Work quickly, but carefully. There are no right or wrong answers.

		Not engaged at all										**Fully engaged**
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1.	*Please rate the extent to which your staff have engaged with the Parent and Carer Engagement core area of the online platform 'The Bubble'. Example modules are structured conversations, parent and carer engagement, welcoming and including families.*											
2.	*Please rate the impact to which your staff have engaged with the Parent and Carer Engagement core area of the online platform 'The Bubble' has had in your school. Example modules are structured conversations, parent and carer engagement, welcoming and including families.*											

SECTION E – ABOUT PROVISION FOR WIDER OUTCOMES AND OPPORTUNITIES IN YOUR SCHOOL

1. Which of the following five wider outcomes have your school chosen to focus upon in AfA?

   - ❑ Attendance
   - ❑ Behaviour and relationships
   - ❑ Pupils' wellbeing: resilience, self-esteem and self-efficacy
   - ❑ Participation - additional opportunities
   - ❑ Managing transitions
   - ❑ None of the above (exclusive)

   [SURVEY TO ONLY SHOW ASSOCIATED TABLES WITH SELECTED OPTIONS ABOVE]

2. Please indicate which of the following activities your school is undertaking in order to improve attendance using AfA funding/resources:

	Improving attendance	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*How effectively does the school work with Parents and Carers to improve attendance of children identified with low attendance?*				
2	*How effective are the strategies currently used to improve the attendance of all pupils?*				

	Behaviour and relationships	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*Are behaviour policies and strategies used consistently across the school and how does the school collect evidence of their impact?*				
2	*Do all students receive equal representation in sanction systems?*				
3	*How effective are approaches and initiatives aimed at improving relationships/friendship skills?*				
4	*How confident are pupils to report bullying and how confident is everyone that the strategies in place to combat bullying are effective and fair?*				

	Pupils' wellbeing: resilience, self-esteem and self-efficacy	**Fully embedded**	**In place**	**Partly in place**	**Not in place**
1	*How effective are the school's strategies to develop and support pupils' well-being and the qualities they need to succeed?*				
2	*Can the school evidence the impact of targeted intervention/provision on learning, well-being, relationships and behaviour?*				
3	*Are all staff aware of the links between speech, language and communication difficulties and behaviour and self-esteem?*				

	Participation - additional opportunities	Fully embedded	In place	Partly in place	Not in place
1	*How well does the school monitor the participation of vulnerable pupils in school life?*				
2	*How effectively are pupils supported in developing positive relationships through participation in specific opportunities?*				
3	*How effectively does the school audit and encourage families' involvement in wider opportunities and partnerships?*				

	Managing transitions	Fully embedded	In place	Partly in place	Not in place
1	*How effectively does the school prepare its pupils, especially those who are vulnerable, for key transition points?*				
2	*Are Parents and Carers involved in and confident about the support provided by the school at key transitions points?*				
3	*How effective are communication systems between settings?*				

3. The next question is related to your staff's engagement with the core area Wider Outcomes and Opportunities of the online platform 'The Bubble'.

   Please answer by clicking on the number that best shows to what extent you have engaged with the specified Bubble module. If you have not engaged, pick a number close to 0. If you have fully engaged, pick a number close to 10. Work quickly, but carefully. There are no right or wrong answers.

		Not engaged at all										Fully engaged
		0	1	2	3	4	5	6	7	8	9	10
1.	*Please rate the extent to which your staff have engaged with the Wider Outcomes and Opportunities core area of the online platform 'The Bubble'. Example modules are developing self-esteem, developing behaviours for attendance, developing behaviours for well-being, etc.*											
2.	*Please rate the impact the Wider Outcomes and Opportunities core area of the online platform 'The Bubble' has had in your school. Example modules are developing self-esteem, developing behaviours for attendance, developing behaviours for well-being, etc.*											

PART 3 – FACTORS AFFECTING IMPLEMENTATION OF AfA

SECTION A – PREPLANNING AND FOUNDATIONS

Please read each statement below and indicate your level of agreement using the scale provided.

		Strongly disagree 1	**Disagree 2**	**Unsure 3**	**Agree 4**	**Strongly Agree 5**
1.	*Staff in my school are open to change.*					
2.	*Staff in my school were involved in the decision to implement Achievement for All.*					
3.	*There are shared expectations for the outcomes of Achievement for All in my school.*					
4.	*There is a high level of awareness of Achievement for All among staff in my school.*					
5.	*Parents of children in my school know about Achievement for All.*					
6.	*My school has the necessary foundations in place for implementing Achievement for All (e.g. existing ethos and practices).*					
7.	*The Achievement for All Champion in my school knew about Achievement for All prior to implementation.*					

SECTION B – IMPLEMENTATION ENVIRONMENT

Please read each statement below and indicate your level of agreement using the scale provided.

		Strongly disagree 1	**Disagree 2**	**Unsure 3**	**Agree 4**	**Strongly Agree 5**
1.	*Our school leadership has been supportive in allowing staff the time to use Achievement for All to develop our practices to support target children*					
2.	*Our school leadership has made Achievement for All a priority.*					
3.	*Our school leadership is committed to implementing Achievement for All.*					
4.	*Achievement for All can be successfully integrated with existing practices in my school and is therefore not a burden.*					
5.	*I can access support within school to help me implement Achievement for All.*					
6.	*There is on-going support and collaboration between staff involved in implementing Achievement for All in my school.*					
7.	*Teachers are enthusiastic about implementing Achievement for All.*					
8.	*The Achievement for All coach in my*					

	school asks for regular updates.					
9.	*It is challenging to allocate time to use Achievement for All to develop our practices to support target children.*					
10.	*It is challenging to allocate time for structured conversations.*					
11.	*Implementing Achievement for All adds a financial burden to the school's cost.*					
12.	*Staff turnover impacts implementation of Achievement for All.*					
13.	*The school size and number of Achievement for All target group children makes implementing Achievement for All difficult.*					
14.	*Starting Achievement for All during the school year (delayed start) has not affected the implementation in my school.*					
15.	*Staff in my school find Achievement for All useful in terms of addressing our school's needs.*					
16.	*Staff in my school find Achievement for All resources and training relevant to improve our target children's outcomes.*					

SECTION C – IMPLEMENTATION SUPPORT SYSTEM

Please read each statement below and indicate your level of agreement using the scale provided.

		Strongly disagree 1	**Disagree 2**	**Unsure 3**	**Agree 4**	**Strongly Agree 5**
1.	*Staff in my school have received sufficient training about Achievement for All.*					
2.	*Staff in my school know where to find help outside of my school to help me implement Achievement for All.*					
3.	*The Achievement for All training our staff received was relevant to their professional needs.*					
4.	*Staff in my school need more training about Achievement for All.*					
5.	*The support we receive from outside of our school helps us to deliver Achievement for All effectively.*					
6.	*Staff in my school know how to access the online platform called the 'Bubble'.*					
7.	*Staff in my school can find suggestions/materials on the online platform the 'Bubble' to support their development.*					
8.	*We need more external support and assistance to help us deliver Achievement for All well.*					
9.	*Our Achievement for All champion has*					

	a good relationship with the school's Achievement for All coach.					
10.	*Staff in my school have had help to reach difficult to engage parents for structured conversations.*					

Many thanks for completing this survey

Appendix H: Histograms for the primary pre-test measures

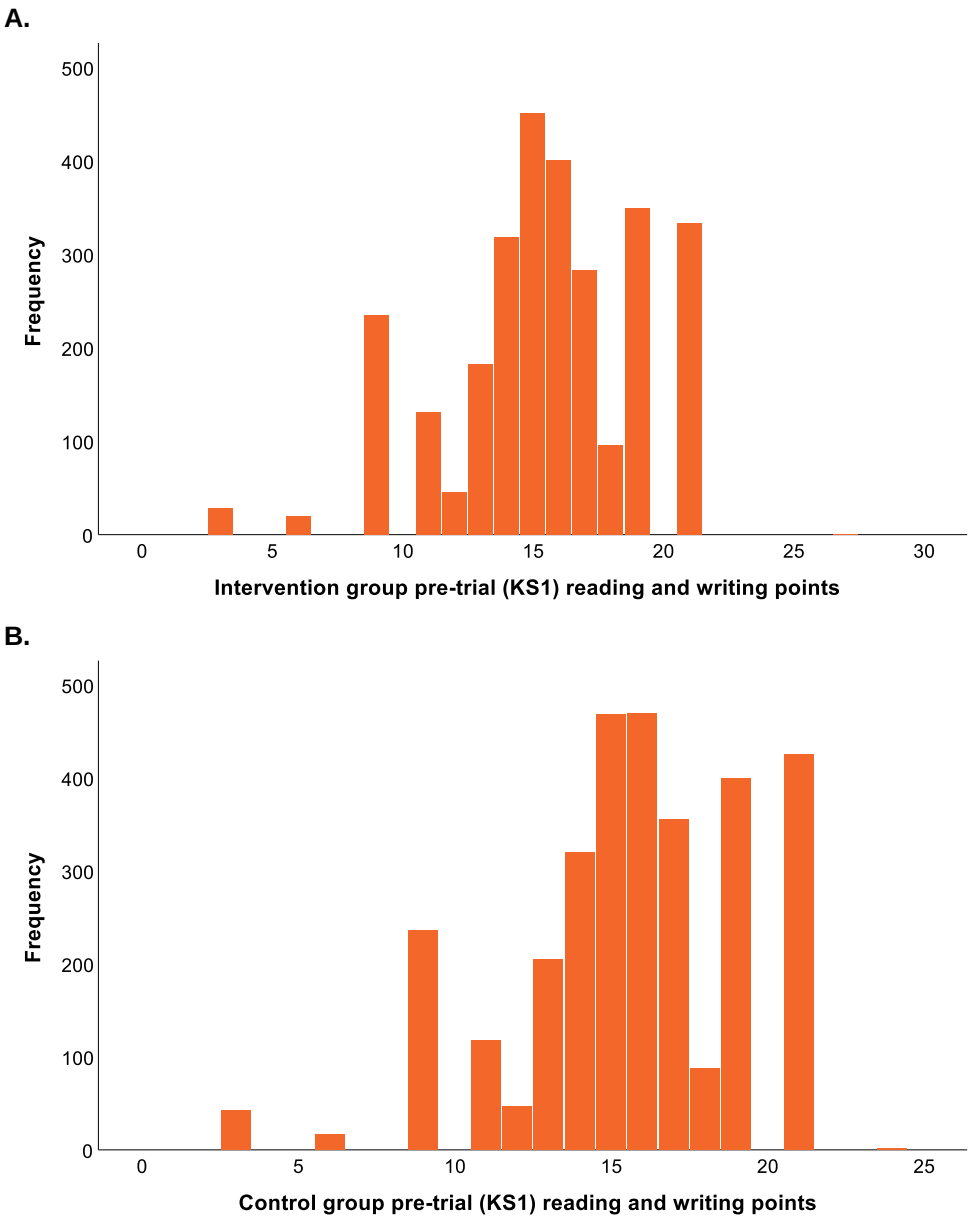


Figure A 1: Histograms of KS1 (pre-trial) reading and writing point scores by trial group. Panels A and B show the distribution of scores for the intervention and control groups, respectively.

Appendix I: MLM ITT and subgroup analyses

Table A1: Reading – Whole Y5 (2016/2017) cohort – complete case

		Empty model β_{0ij} (SE) = 32.633 (0.270)			Baseline model β_{0ij} (SE) = 3.801 (0.464)			Model 1.1 β_{0ij} (SE) = 4.355 (0.503)			Model 1.2 β_{0ij} (SE) = 6.094 (0.964)		
Level		**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil		80.214	1.477	<0.001	42.945	0.806	<0.001	42.948	0.806	<0.001	42.414	0.797	<0.001
	Pre-test (baseline)				1.839	0.026	<0.001	1.838	0.026	<0.001	1.722	0.030	<0.001
	Gender (if male)										0.139	0.175	0.427
	SEND (if eligible)										-2.204	0.271	<0.001
School		7.500	1.174	<0.001	4.634	0.714	<0.001	4.313	0.673	<0.001	3.790	0.606	<0.001
	Trial group (if AfA)							**-1.105**	**0.408**	**0.008**	**-1.142**	**0.388**	**0.004**
	Rand. variables (low as reference): %FSM-Medium										-0.342	0.504	0.498
	Rand. variables (low as reference): %FSM-High										-0.072	0.594	0.904
	Rand. variables (low as reference): %SEND-Medium										0.554	0.487	0.257
	Rand. variables (low as reference): %SEND-High										0.064	0.580	0.913
	Rand. variables (low as reference): %RWM+4-Medium										-0.328	0.487	0.501
	Rand. variables (low as reference): %RWM+4-High										0.986	0.510	0.056
	Usual practice score										0.011	0.061	0.858
		-2*Loglikelihood = 43754.192			-2*Loglikelihood = 38573.398			-2*Loglikelihood = 38566.263			-2*Loglikelihood = 38375.941		
		VPC = 0.086			VPC = 0.097			VPC = 0.091			VPC = 0.082		
		n in model = 6029			*n* in model = 5813			*n* in model = 5813			*n* in model = 5797		

Note. Models were computed in MLwiN using ML estimation.

Model comparisons (chi-square difference test):

- Empty model vs. baseline model: $\Delta X^2 = 5180.794$, $\Delta df = 1$, $p < 0.001$
- Baseline model vs. Model 1.1: $\Delta X^2 = 7.135$, $\Delta df = 2$, $p = 0.028$
- Model 1.1 vs. Model 1.2: $\Delta X^2 = 190.322$, $\Delta df = 9$, $p < 0.001$

Table A2: Reading – Whole Y5 (2016/2017) cohort – MI

		Model 1.1 β_{0ij} (SE) = 5.335 (0.763)			Model 1.2 β_{0ij} (SE) = 7.021 (1.166)		
Level		**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil							
Pre-test (baseline)		1.839	0.026	<0.001	1.727	0.030	<0.001
Gender (if male)					0.095	0.176	0.589
SEND (if eligible)					-2.040	0.266	<0.001
School							
Trial group (if AfA)		**-1.072**	**0.407**	**0.009**	**-1.110**	**0.402**	**0.007**
Rand. variables (low as reference):	%FSM-Medium				-0.349	0.525	0.507
	%FSM-High				-0.066	0.616	0.915
	%SEND-Medium				0.487	0.505	0.337
	%SEND-High				0.047	0.600	0.938
	%RWM+4-Medium				-0.305	0.504	0.546
	%RWM+4-High				1.007	0.529	0.059
Usual practice score					0.007	0.063	0.912

Note. Models were computed with the *lme4* package in R using joint modelling MI (using the *jomo* package).

Table A3: Reading – Y5 (2016/2017) FSM subgroup – complete case

	Empty model β_{0ij} (SE) = 30.134 (0.339)			Baseline model β_{0ij} (SE) = 2.284 (0.818)			Model 2.1 β_{0ij} (SE) = 2.878 (0.859)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil	92.914	3.251	<0.001	50.706	1.793	<0.001	50.758	1.795	<0.001
Pre-test (baseline)				1.905	0.052	<0.001	1.905	0.052	<0.001
School	6.115	1.706	<0.001	5.715	1.284	<0.001	5.211	1.209	<0.001
Trial group (if AfA)							**-1.213**	**0.567**	**0.034**
	-2*Loglikelihood = 12920.070			-2*Loglikelihood = 11692.958			-2*Loglikelihood = 11688.542		
	VPC = 0.062			VPC = 0.101			VPC = 0.093		
	n in model = 1741			*n* in model = 1710			*n* in model = 1710		

Note. Models were computed in MLwin using ML estimation.

Model comparisons (chi-square difference test):

- Empty model vs. baseline model: Δ X^2 = 1227.112, Δ *df* = 1, *p* < 0.001
- Baseline model vs. Model 2.1: Δ X^2 = 4.416, Δ *df* = 1, *p* = 0.036

Table A4: Reading – Y5 (2016/2017) AfA target subgroup – complete case

	Empty model β_{0ij} (SE) = 25.941 (0.463)			Baseline model β_{0ij} (SE) = 3.312 (1.036)			Model 2.2 β_{0ij} (SE) = 4.156 (1.097)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	85.600	3.578	<0.001	60.808	2.587	<0.001	60.895	2.590	<0.001
Pre-test (baseline)				1.745	0.075	<0.001	1.741	0.074	<0.001
School	17.647	3.449	<0.001	9.744	2.115	<0.001	8.896	2.007	<0.001
Trial group (if AfA)							**-1.597**	**0.709**	**0.026**
	-2*Loglikelihood = 9413.399			-2*Loglikelihood = 8663.774			-2*Loglikelihood = 8658.870		
	VPC = 0.171			VPC = 0.138			VPC = 0.127		
	n in model = 1271			*n* in model = 1228			*n* in model = 1228		

Note. Models were computed in MLwiN using ML estimation.

Model comparisons (chi-square difference test):

- Empty model vs. baseline model: Δ X^2 = 749.625, Δ *df* = 1, *p* < 0.001
- Baseline model vs. Model 2.1: Δ X^2 = 4.904, Δ *df* = 1, *p* = 0.027

Table A5: Reading – Y5 (2016/2017) FSM subgroup – MI

	Model 2.1 β_{0ij} (SE) = 4.055 (1.194)		
Level	**β co-efficient**	**SE**	***p***
Pupil			
Pre-test (baseline)	1.898	0.054	<0.001
School			
Trial group (if AfA)	**-1.157**	**0.572**	**0.045**

Note. Models were computed with the *lme4* package in R using joint modelling MI (using the *jomo* package).

Table A6: Reading – Y5 (2016/2017) AfA target subgroup – MI

	Model 2.2 β_{0ij} (SE) = 5.769 (1.482)		
Level	**β co-efficient**	**SE**	***p***
Pupil			
Pre-test (baseline)	1.735	0.075	<0.001
School			
Trial group (if AfA)	**-1.562**	**0.713**	**0.030**

Note. Models were computed with the *lme4* package in R using joint modelling MI (using the *jomo* package).

Table A7: Maths – Whole Y5 (2016/2017) cohort – complete case

		Empty model β_{0ij} (SE) = 74.574 (0.784)			Baseline model β_{0ij} (SE) = -8.151 (1.287)			Model 1.1 β_{0ij} (SE) = -6.787 (1.460)			Model 1.2 β_{0ij} (SE) = -3.375 (3.108)		
Level		**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil		517.023	9.558	<0.001	250.512	4.720	<0.001	250.509	4.720	<0.001	239.779	4.524	<0.001
	Pre-test (baseline)				5.114	0.066	<0.001	5.112	0.066	<0.001	4.636	0.072	<0.001
	Gender (if male)										2.242	0.417	<0.001
	SEND (if eligible)										-9.921	0.627	<0.001
School		67.437	10.001	<0.001	60.223	8.246	<0.001	58.403	8.017	<0.001	53.880	7.460	<0.001
	Trial group (if AfA)							**-2.723**	**1.406**	**0.055**	**-3.036**	**1.357**	**0.027**
	Rand. variables (low as reference): %FSM-Medium										-0.122	1.762	0.945
	Rand. variables (low as reference): %FSM-High										0.95	2.062	0.646
	Rand. variables (low as reference): %SEND-Medium										1.49	1.7	0.383
	Rand. variables (low as reference): %SEND-High										0.602	2.011	0.765
	Rand. variables (low as reference): %RWM+4-Medium										0.038	1.702	0.983
	Rand. variables (low as reference): %RWM+4-High										2.988	1.781	0.096
	Usual practice score										0.303	0.213	0.157
		-2*Loglikelihood = 54623.132			-2*Loglikelihood = 48537.679			-2*Loglikelihood = 48533.977			-2*Loglikelihood = 48143.798		
		VPC = 0.115			VPC = 0.194			VPC = 0.189			VPC = 0.183		
		n in model = 5985			*n* in model = 5768			*n* in model = 5768			*n* in model = 5752		

Note. Models were computed in MLwiN using ML estimation.

Model comparisons (chi-square difference test):

- Empty model vs. baseline model: $\Delta \chi^2 = 6085.453$, $\Delta df = 1$, $p < 0.001$
- Baseline model vs. Model 1.1: $\Delta \chi^2 = 3.136$, $\Delta df = 1$, $p = 0.054$
- Model 1.1 vs. Model 1.2: $\Delta \chi^2 = 390.179$, $\Delta df = 9$, $p < 0.001$

Table A8: Maths – Whole Y5 (2016/2017) cohort – MI

			Model 1.1 β_{0ij} (SE) = -3.800 (2.445)			Model 1.2 β_{0ij} (SE) = -2.509		
Level			**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil								
	Pre-test (baseline)		5.104	0.065	<0.001	4.636	0.070	<0.001
	Gender (if male)					2.138	0.413	<0.001
	SEND (if eligible)					-9.320	0.621	<0.001
School								
	Trial group (if AfA)		**-2.802**	**1.400**	**0.047**	**-3.111**	**1.388**	**0.027**
	Rand. variables (low as reference):	%FSM-Medium				-0.158	1.806	0.931
		%FSM-High				1.021	2.114	0.630
		%SEND-Medium				1.448	1.739	0.407
		%SEND-High				0.598	2.058	0.772
		%RWM+4-Medium				0.048	1.742	0.978
		%RWM+4-High				3.128	1.822	0.089
	Usual practice score					0.317	0.218	0.149

Note. Models were computed with the *lme4* package in R using joint modelling MI (using the *jomo* package).

Table A9: Maths – Y5 (2016/2017) FSM subgroup – complete case

	Empty model β_{0ij} (SE) = 66.687 (1.013)			Baseline model β_{0ij} (SE) = -12.638 (2.172)			Model 2.1 β_{0ij} (SE) = -10.474 (2.347)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	562.429	19.897	<0.001	277.417	9.921	<0.001	277.300	9.916	<0.001
Pre-test (baseline)				5.213	0.130	<0.001	5.207	0.130	<0.001
School	73.536	15.677	<0.001	75.874	13.155	<0.001	71.952	12.636	<0.001
Trial group (if AfA)							**-4.227**	**1.803**	**0.021**
	-2*Loglikelihood = 15834.408			-2*Loglikelihood = 14435.83			-2*Loglikelihood = 14430.424		
	VPC = 0.116			VPC = 0.215			VPC = 0.206		
	n in model = 1714			*n* in model = 1685			*n* in model = 1685		

Note. Models were computed in MLwiN using ML estimation.

Model comparisons (chi-square difference test):

- Empty model vs. baseline model: Δ X^2 = 1398.578, Δ *df* = 1, *p* < 0.001
- Baseline model vs. Model 2.1: Δ X^2 = 5.406, Δ *df* = 1, *p* < 0.020

Table A10: Maths – Y5 (2016/2017) AfA target subgroup – complete case

	Empty model β_{0ij} (SE) = 56.745 (1.261)			Baseline model β_{0ij} (SE) = -10.770 (2.617)			Model 2.2 β_{0ij} (SE) = -8.205 (2.810)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	535.897	22.430	<0.001	325.769	13.879	<0.001	325.860	13.883	<0.001
Pre-test (baseline)				4.874	0.174	<0.001	4.861	0.174	<0.001
School	142.029	25.577	<0.001	91.541	16.501	<0.001	85.479	15.700	<0.001
Trial group (if AfA)							**-4.798**	**1.977**	**0.017**
	-2*Loglikelihood = 11751.538			-2*Loglikelihood = 10758.280			-2*Loglikelihood = 10752.528		
	VPC = 0.210			VPC = 0.219			VPC = 0.208		
	n in model = 1271			*n* in model = 1229			*n* in model = 1229		

Note. Models were computed in MLwiN using ML estimation.

Model comparisons (chi-square difference test):

- Empty model vs. baseline model: Δ X^2 = 993.258, Δ *df* = 1, *p* < 0.001
- Baseline model vs. Model 2.1: Δ X^2 = 5.752, Δ *df* = 1, *p* = 0.017

Table A11: Maths – Y5 (2016/2017) FSM subgroup – MI

	Model 2.1 β_{0ij} (SE) = -5.965 (3.631)		
Level	**β co-efficient**	**SE**	***p***
Pupil			
Pre-test (baseline)	5.184	0.140	<0.001
School			
Trial group (if AfA)	**-4.196**	**1.802**	**0.022**

Note. Models were computed with the *lme4* package in R using joint modelling MI (using the *jomo* package).

Table A12: Maths – Y5 (2016/2017) AfA target subgroup – MI

	Model 2.2 β_{0ij} (SE) = -3.098 (3.937)		
Level	**β co-efficient**	**SE**	***p***
Pupil			
Pre-test (baseline)	4.850	0.168	<0.001
School			
Trial group (if AfA)	**-4.903**	**1.983**	**0.015**

Note. Models were computed with the *lme4* package in R using joint modelling MI (using the *jomo* package).

Table A13: Resilience measure of self-esteem – Whole Y5 (2016/2017) cohort – complete case

		Empty model β_{0ij} (SE) = 11.927 (0.052)			Baseline model β_{0ij} (SE) = 8.657 (0.223)			Model 1.1 β_{0ij} (SE) = 8.669 (0.265)			Model 1.2 β_{0ij} (SE) = 8.450 (0.278)		
Level		**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil		2.982	0.101	<0.001	2.677	0.096	<0.001	2.677	0.096	<0.001	2.617	0.095	<0.001
	Pre-test (baseline)				0.280	0.018	<0.001	0.280	0.018	<0.001	0.264	0.018	<0.001
	Gender (if male)										0.102	0.053	0.054
	SEND (if eligible)										-0.622	0.075	<0.001
School		2.982	0.101	<0.001	0.167	0.051	<0.001	0.167	0.051	0.001	0.141	0.042	0.001
	Trial group (if AfA)							**-0.008**	**0.095**	**0.929**	**-0.010**	**0.091**	**0.915**
	Rand. variables (low as reference): %FSM-Medium										-0.067	0.091	0.461
	Rand. variables (low as reference): %FSM-High										-0.073	0.129	0.573
	Rand. variables (low as reference): %SEND-Medium										0.002	0.093	0.987
	Rand. variables (low as reference): %SEND-High										0.088	0.129	0.495
	Rand. variables (low as reference): %RWM+4-Medium										0.055	0.116	0.632
	Rand. variables (low as reference): %RWM+4-High										0.263	0.126	0.037
	Usual practice score										0.036	0.013	0.006
		-2*Loglikelihood = 18167.058			-2*Loglikelihood = 16555.86			-2*Loglikelihood = 16555.852			-2*Loglikelihood = 16389.600		
		VPC = 0.067			VPC = 0.059			VPC = 0.059			VPC = 0.051		
		n in model = 4589			*n* in model = 4297			*n* in model = 4297			*n* in model = 4282		

Note. Models were computed in MPlus using maximum likelihood estimation with robust standard errors (MLR).

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: $\Delta \chi^2 = 680.836$, $\Delta df = 1$, $p < 0.001$
- Baseline model vs. Model 1.1: $\Delta \chi^2 = 0.008$, $\Delta df = 1$, $p = 0.929$
- Model 1.1 vs. Model 1.2: $\Delta \chi^2 = 171.406$, $\Delta df = 9$, $p < 0.001$

Table A14: Resilience measure of self-esteem – Whole Y5 (2016/2017) cohort - FIML

		Baseline model β_{0ij} (SE) = 8.610 (0.227)			Model 1.1 β_{0ij} (SE) = 8.594 (0.269)			Model 1.2 β_{0ij} (SE) = 8.431 (0.282)		
Level		**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**		
Pupil		2.715	0.101	<0.001	2.715	0.101	<0.001	2.666	0.098	<0.001
Pre-test (baseline)		0.283	0.019	<0.001	0.283	0.019	<0.001	0.267	0.018	<0.001
Gender (if male)								0.074	0.052	0.153
SEND (if eligible)								-0.632	0.077	<0.001
School		0.178	0.053	0.001	0.178	0.053	0.001	0.147	0.044	0.001
Trial group (if AfA)					**0.011**	**0.096**	**0.910**	**0.006**	**0.091**	**0.947**
Rand. variables (low as reference):	%FSM-Medium							-0.094	0.091	0.301
	%FSM-High							-0.079	0.127	0.533
	%SEND-Medium							-0.007	0.091	0.939
	%SEND-High							0.063	0.126	0.618
	%RWM+4-Medium							0.060	0.117	0.607
	%RWM+4-High							0.264	0.128	0.039
Usual practice score								0.035	0.013	0.008
		-2*Loglikelihood = 41582.446			-2*Loglikelihood = 41582.434			-2*Loglikelihood = 56506.382		
		VPC = 0.062			VPC = 0.062			VPC = 0.052		
		n in model = 6109			*n* in model = 6109			*n* in model = 6338		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 1.1: $\Delta \chi^2 = 0.012$, $\Delta df = 1$, $p = 0.913$
- Model 1.1 vs. Model 1.2: $\Delta \chi^2 = -10886.119$, $\Delta df = 13$, $p < 0.001$

Table A15: Resilience measure of self-esteem –Y5 (2016/2017) FSM subgroup – complete case

	Empty model β_{0ij} (SE) = 11.669 (0.074)			Baseline model β_{0ij} (SE) = 9.032 (0.407)			Model 2.1 β_{0ij} (SE) = 9.130 (0.461)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	3.260	0.205	<0.001	2.991	0.200	<0.001	2.992	0.200	<0.001
Pre-test (baseline)				0.229	0.034	<0.001	0.229	0.034	<0.001
School	0.278	0.11	0.011	0.244	0.108	0.024	0.243	0.106	0.022
Trial group (if AfA)							**-0.068**	**0.149**	**0.650**
	-2*Loglikelihood = 5392.186			-2*Loglikelihood = 4870.976			-2*Loglikelihood = 4870.762		
	VPC = 0.079			VPC = 0.075			VPC = 0.075		
	n in model = 1324			*n* in model = 1222			*n* in model = 1222		

Note. Models were computed in MPlus using MLR estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: Δ X^2 = 268.029, Δ *df* = 1, *p* <0.001
- Baseline model vs. model 2.1: Δ X^2 = 0.228, Δ *df* = 1, *p* = 0.633

Table A16: Resilience measure of self-esteem –Y5 (2016/2017) AfA target subgroup – complete case

	Empty model β_{0ij} (SE) = 11.385 (0.079)			Baseline model β_{0ij} (SE) = 8.669 (0.418)			Model 2.2 β_{0ij} (SE) = 8.587 (0.450)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	3.351	0.220	<0.001	3.021	0.211	<0.001	3.021	0.211	<0.001
Pre-test (baseline)				0.241	0.036	<0.001	0.241	0.036	<0.001
School	0.268	0.106	0.012	0.219	0.101	0.030	0.217	0.102	0.032
Trial group (if AfA)							**0.055**	**0.154**	**0.719**
	-2*Loglikelihood = 3957.018			-2*Loglikelihood = 3557.948			-2*Loglikelihood = 3557.816		
	VPC = 0.074			VPC = 0.068			VPC = 0.067		
	n in model = 964			*n* in model = 890			*n* in model = 890		

Note. Models were computed in MPlus using MLR estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: Δ X^2 = 253.926, Δ *df* = 1, *p* <0.001
- Baseline model vs. model 2.2: Δ X^2 = 0.128, Δ *df* = 1, *p* = 0.721

Table A17: Resilience measure of self-esteem – Y5 (2016/2017) FSM subgroup - FIML

	Baseline model β_{0ij} (SE) = 8.954 (0.427)			Model 2.1 β_{0ij} (SE) = 8.938 (0.473)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***
Pupil	3.723	0.193	<0.001	3.080	0.214	<0.001
Pre-test (baseline)	0.236	0.035	<0.001	0.236	0.035	<0.001
School	0.234	0.106	0.027	0.234	0.106	0.234
Trial group (if AfA)				**0.012**	**0.145**	**0.012**
	-2*Loglikelihood = 12269.2			-2*Loglikelihood = 12269.192		
	VPC = 0.059			VPC = 0.071		
	n in model = 1776			*n* in model = 1776		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 2.1: $\Delta X^2 = 0.008$, $\Delta df = 1$, $p = 0.929$

Table A18: Resilience measure of self-esteem – Y5 (2016/2017) AfA target subgroup - FIML

	Baseline model β_{0ij} (SE) = 8.539 (0.455)			Model 2.1 β_{0ij} (SE) = 8.421 (0.492)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***
Pupil	3.151	0.219	<0.001	3.153	0.220	<0.001
Pre-test (baseline)	0.252	0.039	<0.001	0.252	0.039	<0.001
School	0.213	0.100	<0.001	0.209	0.100	0.037
Trial group (if AfA)				**0.081**	**0.151**	**0.592**
	-2*Loglikelihood = 9092.506			-2*Loglikelihood = 9092.216		
	VPC = 0.063			VPC = 0.062		
	n in model = 1317			*n* in model = 1317		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 2.1: $\Delta X^2 = 0.280$, $\Delta df = 1$, $p = 0.597$

Table A19: Resilience measure of goals and aspirations – Whole Y5 (2016/2017) cohort – complete case

Level			Empty model β_{0ij} (SE) = 8.361 (0.043)			Baseline model β_{0ij} (SE) = 6.370 (0.144)			Model 1.1 β_{0ij} (SE) = 6.262 (0.178)			Model 1.2 β_{0ij} (SE) = 5.767 (0.227)		
			β co-efficient	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil			2.669	0.088	<0.001	2.485	0.089	<0.001	2.484	0.089	<0.001	2.481	0.089	<0.001
	Pre-test (baseline)					0.243	0.016	<0.001	0.242	0.016	<0.001	0.238	0.016	<0.001
	Gender (if male)											0.069	0.05	0.168
	SEND (if eligible)											-0.247	0.068	<0.001
School			0.131	0.027	<0.001	0.106	0.025	<0.001	0.104	0.024	<0.001	0.078	0.019	<0.001
	Trial group (if AfA)								**0.075**	**0.080**	**0.353**	**0.072**	**0.074**	**0.334**
	Rand. variables (low as reference):	%FSM-Medium										0.061	0.087	0.486
		%FSM-High										0.187	0.107	0.079
		%SEND-Medium										-0.001	0.094	0.992
		%SEND-High										0.076	0.105	0.470
		%RWM+4-Medium										0.012	0.090	0.893
		%RWM+4-High										0.113	0.100	0.260
	Usual practice score											0.039	0.012	0.001
			-2*Loglikelihood = 17621.33			-2*Loglikelihood = 16132.638			-2*Loglikelihood = 16131.768			-2*Loglikelihood = 16058.894		
			VPC = 0.047			VPC = 0.041			VPC = 0.040			VPC = 0.030		
			n in model = 4583			*n* in model = 4277			*n* in model = 4277			*n* in model = 4264		

Note. Models were computed in MPlus using MLR estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: $\Delta \chi^2 = 895.831$, $\Delta df = 1$, $p < 0.001$
- Baseline model vs. Model 1.1: $\Delta \chi^2 = 0.904$, $\Delta df = 1$, $p = 0.342$
- Model 1.1 vs. Model 1.2: $\Delta \chi^2 = 75.113$, $\Delta df = 9$, $p < 0.001$

Table A20: Resilience measure of goals and aspirations– Whole Y5 (2016/2017) cohort – FIML

		Baseline model β_{0ij} (SE) = 6.350 (0.174)			Model 1.1 β_{0ij} (SE) = 8.238 (0.036)			Model 1.2 β_{0ij} (SE) = 5.755 (0.226)		
Level		**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**		
Pupil		2.938	0.074	<0.001	2.512	0.086	<0.001	2.505	0.085	<0.001
Pre-test (baseline)		0.244	0.017	<0.001	0.244	0.017	<0.001	0.238	0.016	<0.001
Gender (if male)								0.058	0.050	0.243
SEND (if eligible)								-0.280	0.068	<0.001
School		0.106	0.025	<0.001	0.104	0.024	<0.001	0.077	0.019	<0.001
Trial group (if AfA)					**0.078**	**0.079**	**0.362**	**0.078**	**0.073**	**0.283**
Rand. variables (low as reference):	%FSM-Medium							0.031	0.086	0.724
	%FSM-High							0.179	0.103	0.083
	%SEND-Medium							0.015	0.092	0.872
	%SEND-High							0.063	0.105	0.546
	%RWM+4-Medium							-0.009	0.089	0.923
	%RWM+4-High							0.097	0.099	0.330
Usual practice score								0.042	0.012	<0.001
		-2*Loglikelihood = 40018.342			-2*Loglikelihood = 40017.386			-2*Loglikelihood = 55006.166		
		VPC = 0.035			VPC = 0.040			VPC = 0.030		
		n in model = 6094			*n* in model = 6094			*n* in model = 6338		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 1.1: $\Delta \chi^2 = 0.998$, $\Delta df = 1$, $p = 0.318$
- Model 1.1 vs. model 2.1: $\Delta \chi^2 = -10828.900$, $\Delta df = 13$, $p < 0.001$

Table A21: Resilience measure of goals and aspirations –Y5 (2016/2017) FSM subgroup – complete case

	Empty model β_{0ij} (SE) = 8.288 (0.067)			Baseline model β_{0ij} (SE) = 6.114 (0.261)			Model 2.1 β_{0ij} (SE) = 6.101 (0.300)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	3.009	0.182	<0.001	2.832	0.190	<0.001	2.832	0.189	<0.001
Pre-test (baseline)				0.264	0.030	<0.001	0.264	0.030	<0.001
School	0.177	0.055	0.001	0.127	0.047	0.007	0.126	0.048	0.008
Trial group (if AfA)							**0.010**	**0.128**	**0.940**
	-2*Loglikelihood = 5277.792			-2*Loglikelihood = 4785.144			-2*Loglikelihood = 4785.138		
	VPC = 0.056			VPC = 0.043			VPC = 0.043		
	n in model = 1326			*n* in model = 1223			*n* in model = 1223		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: Δ x^2 = 344.389, Δ *df* = 1, *p* <0.001
- Baseline model vs. model 2.1: Δ x^2 = 0.006, Δ *df* = 1, *p* = 0.938

Table A22: Resilience measure of goals and aspirations –Y5 (2016/2017) AfA target subgroup – complete case

	Empty model β_{0ij} (SE) = 8.083 (0.064)			Baseline model β_{0ij} (SE) = 6.211 (0.311)			Model 2.2 β_{0ij} (SE) = 6.169 (0.347)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	3.355	0.178	<0.001	3.253	0.186	<0.001	3.235	0.186	<0.001
Pre-test (baseline)				0.235	0.038	<0.001	0.235	0.038	<0.001
School	0.045	0.052	0.386	0.025	0.043	0.555	0.025	0.042	0.552
Trial group (if AfA)							**0.030**	**0.125**	**0.811**
	-2*Loglikelihood = 3910.704			-2*Loglikelihood = 3573.312			-2*Loglikelihood = 3573.256		
	VPC = 0.013			VPC = 0.008			VPC = 0.008		
	n in model = 963			*n* in model = 889			*n* in model = 889		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: Δ x^2 = 314.116, Δ *df* = 1, *p* <0.001
- Baseline model vs. model 2.2: Δ x^2 = 0.058, Δ *df* = 1, *p* = 0.810

Table A23: Resilience measure of goals and aspirations – Y5 (2016/2017) FSM subgroup – FIML

	Baseline model β_{0ij} (SE) = 6.150 (0.257)			Model 2.1 β_{0ij} (SE) = 6.141 (0.303)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***
Pupil	2.820	0.180	<0.001	2.820	0.179	<0.001
Pre-test (baseline)	0.261	0.029	<0.001	0.262	0.029	<0.001
School	0.130	0.045	0.004	0.130	0.046	0.005
Trial group (if AfA)				**0.005**	**0.126**	**0.971**
	-2*Loglikelihood = 11905.784			-2*Loglikelihood = 11905.782		
	VPC = 0.044			VPC = 0.044		
	n in model = 1769			*n* in model = 1769		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 2.1: $\Delta X^2 = 0.002$, $\Delta df = 1$, $p = 0.964$

Table A24: Resilience measure of goals and aspirations – Y5 (2016/2017) AfA target subgroup – FIML

	Baseline model β_{0ij} (SE) = 6.231 (0.041)			Model 2.2 β_{0ij} (SE) = 6.116 (0.350)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***
Pupil	3.203	0.038	<0.001	3.204	0.173	<0.001
Pre-test (baseline)	0.232	0.038	<0.001	0.231	0.038	<0.001
School	0.025	0.041	0.536	0.023	0.039	0.546
Trial group (if AfA)				**0.081**	**0.120**	**0.498**
	-2*Loglikelihood = 8855.378			-2*Loglikelihood = 8854.926		
	VPC = 0.008			VPC = 0.007		
	n in model = 1308			*n* in model = 1308		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 2.2: $\Delta X^2 = 0.481$, $\Delta df = 1$, $p = 0.488$

Table A25: Resilience measure of family connection – Whole Y5 (2016/2017) cohort – complete case

Level			Empty model β_{0ij} (SE) = 18.201 (0.062)			Baseline model β_{0ij} (SE) = 13.960 (0.389)			Model 1.1 β_{0ij} (SE) = 13.772 (0.448)			Model 1.2 β_{0ij} (SE) = 13.801 (0.451)		
			β co-efficient	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil			4.239	0.220	<0.001	3.876	0.183	<0.001	3.867	0.183	<0.001	3.822	0.182	<0.001
	Pre-test (baseline)					0.237	0.020	<0.001	0.237	0.020	<0.001	0.224	0.020	<0.001
	Gender (if male)											-0.332	0.062	<0.001
	SEND (if eligible)											-0.369	0.119	0.002
School			0.301	0.074	<0.001	0.230	0.054	<0.001	0.225	0.053	<0.001	0.203	0.054	<0.001
	Trial group (if AfA)								**0.129**	**0.111**	**0.248**	**0.118**	**0.109**	**0.279**
	Rand. variables (low as reference):	%FSM-Medium										-0.024	0.120	0.843
		%FSM-High										-0.028	0.156	0.855
		%SEND-Medium										0.096	0.120	0.427
		%SEND-High										0.170	0.153	0.264
		%RWM+4-Medium										-0.078	0.136	0.563
		%RWM+4-High										0.011	0.154	0.944
	Usual practice score											0.040	0.016	0.013
			-2*Loglikelihood = 19112.310			-2*Loglikelihood = 17071.776			-2*Loglikelihood = 17070.474			-2*Loglikelihood = 16965.536		
			VPC = 0.066			VPC = 0.056			VPC = 0.055			VPC = 0.050		
			n in model = 4430			*n* in model = 4045			*n* in model = 4045			*n* in model = 4033		

Note. Models were computed in MPlus using MLR estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: $\Delta \chi^2 = 3211.416$, $\Delta df = 1$, $p < 0.001$
- Baseline model vs. model 1.1: $\Delta \chi^2 = 1.349$, $\Delta df = 1$, $p = 0.246$
- Model 1.1 vs. model 1.2: $\Delta \chi^2 = 96.727$, $\Delta df = 9$, $p < 0.001$

Table A26: Resilience measure of family connection – Whole Y5 (2016/2017) cohort – FIML

		Baseline model β_{0ij} (SE) = 13.914 (0.410)			Model 1.1 β_{0ij} (SE) = 13.698 (0.470)			Model 1.2 β_{0ij} (SE) = 13.769 (0.457)		
Level		**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil		6.405	0.286	<0.001	6.405	0.286	<0.001	3.862	0.192	<0.001
Pre-test (baseline)		0.239	0.021	<0.001	0.239	0.021	<0.001	0.225	0.021	<0.001
Gender (if male)								-0.328	0.061	<0.001
SEND (if eligible)								-0.406	0.114	<0.001
School		0.244	0.052	<0.001	0.236	0.050	<0.001	0.220	0.052	<0.001
Trial group (if AfA)					**0.148**	**0.112**	**0.185**	**0.144**	**0.11**	**0.189**
Rand. variables (low as reference):	%FSM-Medium							-0.017	0.121	0.886
	%FSM-High							-0.032	0.154	0.836
	%SEND-Medium							0.107	0.121	0.376
	%SEND-High							0.170	0.154	0.270
	%RWM+4-Medium							-0.025	0.14	0.857
	%RWM+4-High							0.049	0.155	0.751
Usual practice score								0.034	0.016	0.037
		-2*Loglikelihood = 45314.698			-2*Loglikelihood = 45312.970			-2*Loglikelihood = 60271.254		
		VPC = 0.037			VPC = 0.036			VPC = 0.054		
		n in model = 6037			*n* in model = 6037			*n* in model = 6338		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 1.1: $\Delta \chi^2 = 1.820$, $\Delta df = 1$, $p = 0.177$
- Model 1.1 vs. Model 1.2: $\Delta \chi^2 = -10356.440$, $\Delta df = 13$, $p < 0.001$

Table A27: Resilience measure of family connection – Y5 (2016/2017) FSM subgroup – complete case

	Empty model β_{0ij} (SE) = 17.991 (0.100)			Baseline model β_{0ij} (SE) = 14.065 (0.634)			Model 2.1 β_{0ij} (SE) = 13.939 (0.734)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	4.861	0.413	<0.001	4.400	0.329	<0.001	4.400	0.329	<0.001
Pre-test (baseline)				0.222	0.033	<0.001	0.221	0.033	<0.001
School	0.528	0.183	0.004	0.412	0.145	0.005	0.409	0.146	0.005
Trial group (if AfA)							**0.088**	**0.186**	**0.636**
	-2*Loglikelihood = 5741.472			-2*Loglikelihood = 5087.472			-2*Loglikelihood = 5087.250		
	VPC = 0.098			VPC = 0.086			VPC = 0.085		
	n in model = 1281			*n* in model = 1162			*n* in model = 1162		

Note. Models were computed in MPlus using MLR estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: Δ x^2 = 2212.449, Δ *df* = 1, *p* <0.001
- Baseline model vs. model 2.1: Δ x^2 = 0.220, Δ *df* = 1, *p* = 0.639

Table A28: Resilience measure of family connection – Y5 (2016/2017) AfA target subgroup – complete case

	Empty model β_{0ij} (SE) = 17.915 (0.106)			Baseline model β_{0ij} (SE) = 14.732 (0.723)			Model 2.2 β_{0ij} (SE) = 14.831 (0.831)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	5.004	0.473	<0.001	4.756	0.434	<0.001	4.754	0.433	<0.001
Pre-test (baseline)				0.183	0.039	<0.001	0.183	0.039	<0.001
School	0.522	0.268	0.051	0.327	0.165	0.048	0.331	0.163	0.043
Trial group (if AfA)							**-0.068**	**0.187**	**0.715**
	-2*Loglikelihood = 4216.074			-2*Loglikelihood = 3751.332			-2*Loglikelihood = 3751.208		
	VPC = 0.094			VPC = 0.064			VPC = 0.065		
	n in model = 933			*n* in model = 843			*n* in model = 843		

Note. Models were computed in MPlus using MLR estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: Δ x^2 = 1085.339, Δ *df* = 1, *p* <0.001
- Baseline model vs. model 2.2: Δ x^2 = 0.131, Δ *df* = 1, *p* = 0.717

Table A29: Resilience measure of family connection – Y5 (2016/2017) FSM subgroup – FIML

	Baseline model β_{0ij} (SE) = 13.942 (0.685)			Model 2.1 β_{0ij} (SE) = 13.775 (0.788)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***
Pupil	4.534	0.368	<0.001	4.535	0.369	<0.001
Pre-test (baseline)	0.229	0.036	<0.001	0.228	0.036	<0.001
School	0.426	0.130	0.001	0.419	0.13	0.001
Trial group (if AfA)				**0.117**	**0.185**	**0.527**
	-2*Loglikelihood = 13608.816			-2*Loglikelihood = 13608.420		
	VPC = 0.086			VPC = 0.085		
	n in model = 1757			*n* in model = 1757		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 2.1: $\Delta \chi^2 = 0.391$, $\Delta df = 1$, $p = 0.532$

Table A30: Resilience measure of family connection – Y5 (2016/2017) AfA target subgroup – FIML

	Baseline model β_{0ij} (SE) = 14.677 (0.765)			Model 2.2 β_{0ij} (SE) = 14.790 (0.895)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***
Pupil	4.802	0.444	<0.001	4.799	0.444	<0.001
Pre-test (baseline)	0.185	0.041	<0.001	0.186	0.041	<0.001
School	0.420	0.179	0.019	0.423	0.177	0.017
Trial group (if AfA)				**-0.078**	**0.191**	**0.685**
	-2*Loglikelihood = 10049.468			-2*Loglikelihood = 10049.314		
	VPC = 0.080			VPC = 0.081		
	n in model = 1294			*n* in model = 1294		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 2.2: $\Delta \chi^2 = 0.160$, $\Delta df = 1$, $p = 0.689$

Table A31: Resilience measure of school connection – Whole Y5 (2016/2017) cohort – complete case

		Empty model β_{0ij} (SE) = 17.280 (0.100)			Baseline model β_{0ij} (SE) = 12.429 (0.385)			Model 1.1 β_{0ij} (SE) = 11.749 (0.478)			Model 1.2 β_{0ij} (SE) = 11.351 (0.567)		
Level		**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil		8.329	0.424	<0.001	7.564	0.375	<0.001	7.564	0.375	<0.001	7.507	0.369	<0.001
	Pre-test (baseline)				0.281	0.020	<0.001	0.281	0.020	<0.001	0.276	0.020	<0.001
	Gender (if male)										-0.473	0.088	<0.001
	SEND (if eligible)										0.146	0.131	0.267
School		0.868	0.137	<0.001	0.742	0.130	<0.001	0.692	0.129	<0.001	0.609	0.114	<0.001
	Trial group (if AfA)							**0.454**	**0.183**	**0.013**	**0.455**	**0.177**	**0.010**
	Rand. variables (low as reference): %FSM-Medium										0.347	0.199	0.081
	Rand. variables (low as reference): %FSM-High										0.636	0.248	0.010
	Rand. variables (low as reference): %SEND-Medium										0.227	0.205	0.269
	Rand. variables (low as reference): %SEND-High										-0.138	0.252	0.583
	Rand. variables (low as reference): %RWM+4-Medium										-0.133	0.227	0.558
	Rand. variables (low as reference): %RWM+4-High										0.034	0.234	0.886
	Usual practice score										0.039	0.029	0.182
		-2*Loglikelihood = 22444.430			-2*Loglikelihood = 20078.542			-2*Loglikelihood = 20072.578			-2*Loglikelihood = 19968.948		
		VPC = 0.094			VPC = 0.089			VPC = 0.084			VPC = 0.075		
		n in model = 4492			*n* in model = 4097			*n* in model = 4097			*n* in model = 4084		

Note. Models were computed in MPlus using MLR estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: $\Delta x^2 = 2319.953$, $\Delta df = 1$, $p < 0.001$
- Baseline model vs. model 1.1: $\Delta x^2 = 5.464$, $\Delta df = 1$, $p = 0.019$
- Model 1.1 vs. model 1.2: $\Delta x^2 = 104.468$, $\Delta df = 9$, $p < 0.001$

Table A32: Resilience measure of school connection – Whole Y5 (2016/2017) cohort – FIML

		Baseline model β_{0ij} (SE) = 12.357 (0.394)			Model 1.1 β_{0ij} (SE) = 11.703 (0.485)			Model 1.2 β_{0ij} (SE) = 11.314 (0.567)		
Level		**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil		7.702	0.391	<0.001	7.702	0.391	<0.001	7.648	0.385	<0.001
Pre-test (baseline)		0.285	0.021	<0.001	0.285	0.021	<0.001	0.277	0.021	<0.001
Gender (if male)								-0.487	0.087	<0.001
SEND (if eligible)								0.121	0.123	0.324
School		0.736	0.127	<0.001	0.689	0.126	<0.001	0.607	0.112	<0.001
Trial group (if AfA)					**0.463**	**0.181**	**0.016**	**0.438**	**0.175**	**0.012**
Rand. variables (low as reference):	%FSM-Medium							0.314	0.195	0.108
	%FSM-High							0.604	0.227	0.008
	%SEND-Medium							0.244	0.197	0.214
	%SEND-High							-0.083	0.238	0.727
	%RWM+4-Medium							-0.147	0.226	0.517
	%RWM+4-High							0.054	0.230	0.815
Usual practice score								0.042	0.029	0.147
		-2*Loglikelihood = 50242.16			-2*Loglikelihood = 50236.522			-2*Loglikelihood = 65216.06		
		VPC = 0.087			VPC = 0.082			VPC = 0.074		
		n in model = 6054			*n* in model = 6054			*n* in model = 6338		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 1.1: $\Delta X^2 = 5.242$, $\Delta df = 1$, $p = 0.022$
- Model 1.1 vs. model 1.2: $\Delta X^2 = -10883.007$, $\Delta df = 13$, $p < 0.001$

Table A33: Resilience measure of school connection – Y5 (2016/2017) FSM subgroup – complete case

	Empty model β_{0ij} (SE) = 17.272 (0.133)			Baseline model β_{0ij} (SE) = 12.760 (0.693)			Model 2.1 β_{0ij} (SE) = 12.185 (0.777)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	8.618	0.689	<0.001	7.813	0.685	<0.001	7.809	0.684	<0.001
Pre-test (baseline)				0.261	0.037	<0.001	0.261	0.037	<0.001
School	0.950	0.283	0.001	0.914	0.281	0.001	0.888	0.276	0.001
Trial group (if AfA)							**0.382**	**0.260**	**0.143**
	-2*Loglikelihood = 6530.962			-2*Loglikelihood = 5780.41			-2*Loglikelihood = 5778.308		
	VPC = 0.099			VPC = 0.105			VPC = 0.102		
	n in model = 1292			*n* in model = 1165			*n* in model = 1165		

Note. Models were computed in MPlus using MLR estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: Δ X^2 = 388.666, Δ *df* = 1, *p* <0.001
- Baseline model vs. model 2.1: Δ X^2 = 2.028, Δ *df* = 1, *p* = 0.154

Table A34: Resilience measure of school connection – Y5 (2016/2017) AfA target subgroup – complete case

	Empty model β_{0ij} (SE) = 17.348 (0.132)			Baseline model β_{0ij} (SE) = 13.568 (0.664)			Model 2.2 β_{0ij} (SE) = 12.928 (0.732)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***	**β co-efficient**	**SE**	***p***
Pupil	7.237	0.604	<0.001	6.604	0.547	<0.001	6.613	0.547	<0.001
Pre-test (baseline)				0.221	0.036	<0.001	0.221	0.036	<0.001
School	0.959	0.287	0.001	0.787	0.257	0.002	0.725	0.239	0.002
Trial group (if AfA)							**0.425**	**0.247**	**0.085**
	-2*Loglikelihood = 4587.098			-2*Loglikelihood = 4051.074			-2*Loglikelihood = 4048.266		
	VPC = 0.117			VPC = 0.106			VPC = 0.099		
	n in model = 936			*n* in model = 843			*n* in model = 843		

Note. Models were computed in MPlus using MLR estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Empty model vs. baseline model: Δ X^2 = 616.971, Δ *df* = 1, *p* <0.001
- Baseline model vs. model 2.2: Δ X^2 = 3.137, Δ *df* = 1, *p* = 0.077

Table A35: Resilience measure of school connection – Y5 (2016/2017) FSM subgroup – FIML

	Baseline model β_{0ij} (SE) = 12.635 (0.705)			Model 2.1 β_{0ij} (SE) = 12.096 (0.787)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***
Pupil	8.051	0.673	<0.001	8.048	0.672	<0.001
Pre-test (baseline)	0.268	0.038	<0.001	0.268	0.038	<0.001
School	0.780	0.243	0.001	0.756	0.239	0.002
Trial group (if AfA)				**0.359**	**0.246**	**0.144**
	-2*Loglikelihood = 14682.822			-2*Loglikelihood = 14680.726		
	VPC = 0.088			VPC = 0.086		
	n in model = 1758			*n* in model = 1758		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 2.1: Δ X^2 = 2.003, Δ *df* = 1, *p* = 0.157

Table A36: Resilience measure of school connection – Y5 (2016/2017) AfA target subgroup – FIML

	Baseline model β_{0ij} (SE) = 13.424 (0.697)			Model 2.2 β_{0ij} (SE) = 12.978 (0.826)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***
Pupil	6.806	0.559	<0.001	6.812	0.559	<0.001
Pre-test (baseline)	0.229	0.038	<0.001	0.230	0.038	<0.001
School	0.898	0.254	<0.001	0.862	0.244	<0.001
Trial group (if AfA)				**0.295**	**0.253**	**0.244**
	-2*Loglikelihood = 10628.884			-2*Loglikelihood = 10627.578		
	VPC = 0.117			VPC = 0.112		
	n in model = 1302			*n* in model = 1302		

Note. Models were computed in MPlus using FIML estimation.

Model comparisons (chi-square difference test using the scaling correction factor):

- Baseline model vs. model 2.2: Δ X^2 = 1.399, Δ *df* = 1, *p* = 0.237

Table A37: RQ3 models –Y5 (2016/2017) intervention schools only

	Empty model β_{0ij} (SE) = 31.960 (0.339)			Baseline model β_{0ij} (SE) = 2.482 (0.675)			Model 1.2 β_{0ij} (SE) = 2.058 (1.822)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***p***
Pupil	85.111	2.266	<0.001	44.922	1.223	<0.001	44.326	1.464	<0.001
Pre-test (baseline)				1.887	0.039	<0.001	1.908	0.047	<0.001
FSM (if eligible)									
AfA target group (if member)									
School	5.150	1.298	<0.001	4.519	1.016	<0.001	3.527	1.028	0.002
Leadership & inclusion							-0.726	0.341	0.039
Teaching & learning							0.564	0.375	0.141
Provision & wider outcomes							-0.168	0.352	0.636
SC fidelity							0.784	0.709	0.276
SC dosage							-0.644	0.547	0.247
	-2*Loglikelihood = 21095.594			-2*Loglikelihood = 18462.997			-2*Loglikelihood = 12504.094		
	VPC = 0.057			VPC = 0.091			VPC = 0.074		

Note. Models were computed in MLwiN using ML estimation. SC = structured conversations.

Table A38: RQ3 models –Y5 (2016/2017) intervention schools only – with FSM and AfA target interactions

	FSM interaction model β_{0ij} (SE) = 3.705 (1.882)			AfA target interaction model β_{0ij} (SE) = 5.633 (1.880)		
Level	**β co-efficient**	**SE**	***p***	**β co-efficient**	**SE**	***P***
Pupil	44.042	1.454	<0.001	42.990	1.420	<0.001
Pre-test (baseline)	1.877	0.048	<0.001	1.760	0.051	<0.001
FSM (if eligible)	-2.517	1.783	0.158			
AfA target group (if member)				-7.352	1.977	<0.001
School	3.001	0.909	0.002	3.359	0.983	0.002
Leadership & inclusion	-0.599	0.338	0.084	-0.69	0.343	0.051
Teaching & learning	0.381	0.374	0.315	0.456	0.379	0.236
Provision & wider outcomes	-0.039	0.359	0.914	-0.062	0.356	0.863
SC fidelity	0.579	0.715	0.423	0.501	0.713	0.486
SC dosage	-0.816	0.545	0.143	-0.674	0.551	0.229
FSM interactions						
FSM*leadership & inclusion	-0.382	0.331	0.249			
FSM*teaching & learning	0.527	0.38	0.166			
FSM*provision & wider outcomes	-0.37	0.354	0.296			
FSM*SC fidelity	0.215	0.752	0.775			
FSM*SC dosage	0.452	0.591	0.444			
AfA target interactions						
AfA*leadership & inclusion				0.152	0.363	0.675
AfA*teaching & learning				0.491	0.413	0.235
AfA*provision & wider outcomes				-0.46	0.373	0.218
AfA*SC fidelity				1.538	0.804	0.056
AfA*SC dosage				0.466	0.621	0.453
	-2*Loglikelihood = 12487.229			-2*Loglikelihood = 12446.098		
	VPC = 0.064			VPC = 0.072		

Note. Models were computed in MLwiN using ML estimation. SC = structured conversations.

Appendix J: Multi-level negative binomial regression models for the attendance data

Multi-level negative binomial regression models with complete case data for the Y5 (2016/2017) AfA target subgroup. All models were estimated in MLwiN (Version 2.22) and the outcome measure in each case was attendance[31].

Empty model with just random intercept:

- Fixed β_{0j} (SE) = 1.744 (0.080), p<0.001; π_{ij} = 5.720
- Random μ_{0j} (SE) = 0.837 (0.105), p<0.001

Baseline model with just pre-trial count entered as a covariate:

- Fixed β_{0j} (SE) = 1.253 (0.078), p<0.001; π_{ij} = 3.501
- Random μ_{0j} (SE) = 0.254 (0.085), p=0.003
- Fixed *pre-trial absence* (pupil level) (SE) = 0.057 (0.005), p<0.001

Model 1.1 - primary analysis:

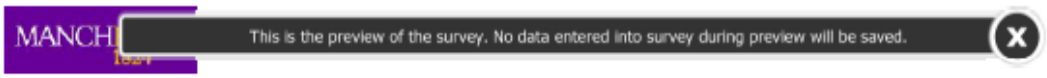
- Fixed β_{0j} (SE) = 1.280 (0.105), p<0.001; π_{ij} = 3.597
- Random μ_{0j} (SE) = 0.242 (0.083), p=0.004
- *Pre-trial absence* (pupil level) (SE) = 0.058 (0.005), p<0.001
- ***Group (school level)* (SE) = -0.063 (0.146), p=0.666**

Model 1.2 - sensitivity analysis (with all additional covariates entered):

- Fixed β_{0j} (SE) = 0.728 (0.254), p=0.005
- Random μ_{0j} (SE) = 0.055 (0.051), p=0.283
- *Pre-trial absence* (pupil level) (SE) = 0.060 (0.005), p<0.001
- ***Group (school level)* (SE) = -0.019 (0.117), p=0.872**
- *SEND* (pupil level) (SE) = 0.139 (0.113), p=0.221
- *Gender* (pupil level) (SE) = 0.083 (0.109), p=0.448
- *FSM-medium* (school level) (SE) = 0.430 (0.150), p=0.005
- *FSM-high* (school level) (SE) = 0.826 (0.181), p<0.001
- *SEND-medium* (school level) (SE) = 0.035 (0.147), p=0.812
- *SEND-high* (school level) (SE) = -0.168 (0.178), p=0.347
- *RWM-medium* (school level) (SE) = -0.247 (0.147), p=0.955
- *RWM-high* (school level) (SE) = 0.015 (0.153), p=0.922
- *Usual practice score* (school level) (SE) = 0.005 (0.018), p=0.782

[31] π_{ij} denotes exp(β_{0j}) (i.e. the estimated mean of post-trial unauthorised absences) and μ_{0j} denotes the level 2 variance (i.e. the between school variance of post-trial unauthorised absences)

Appendix K: The Usual Practice Survey



Achievement for All - AfA Trial (RCT Strand)

Section 1 - About you and your role(s) in school

1. Please indicate which of the following roles you undertake in your school (tick any that apply):

- ☐ Head teacher
- ☐ Deputy or assistant head teacher
- ☐ Special educational needs co-ordinator/inclusion manager
- ☐ Head of year or key stage
- ☐ Class or subject teacher
- ☐ Other (please specify)

2. Are you part of your school's leadership/management team?

- ○ Yes
- ○ No

Section 2 - Continuing Professional Development (CPD)

The following questions are about the nature and type of Continuing Professional Development (CPD) undertaken in your school since September 2014.

1. Has CPD been undertaken in the following areas in your school since September 2014?

- Please note that the descriptions are only examples of what you may have covered under each category of the CPD outlined below. For example, if you have received CPD for 'coaching for inclusive leadership', but the description does not match or apply, you may still select yes. But enter your own description (in no more than 25 words).
- If you have answered yes, please rate on a scale of 1 to 5 how extensive the CPD was (with 5 being most extensive and 1 least extensive). For example, most extensive could involve an intensive, comprehensive whole-school CPD, etc., whereas least extensive could involve a limited number of staff or a one off CPD session, etc.
- If you have answered yes, please also make any additional notes under 'briefly describe' (if required).

a. Leade

This is the preview of the survey. No data entered into survey during preview will be saved.

	Yes/No	If yes, how extensive?	If yes, please briefly describe:
Coaching for inclusive leadership Coaching involves working with school leaders to focus on inclusion with the aim of improving aspirations, access and achievement of pupils with SEND.			
Collaborative action research Collaborative action research is a process of inquiry into practice that is undertaken with others as a means to work towards an agreed aim.			
Effective use of Pupil Premium Exploring ways to use Pupil Premium funding more effectively in order to improve provision and outcomes for vulnerable and disadvantaged pupils.			
Governance for inclusion Development of school governance constitution and processes in order to improve academic and wider outcomes for vulnerable individuals and groups.			
Networking based on need Working with other schools with similar contexts and needs in order to learn from one another and develop more effective provision for vulnerable pupils.			
Maximising the impact of Teaching and Learning Assistants Developing practices that can enhance pupils' learning through more effective use and deployment of teaching and learning assistants.			
Development of middle leaders Supporting aspiring and serving middle leaders (e.g. SENCO) in the school to develop provision in their designated area of responsibility.			
Other Please complete, if you had any other CPD pertaining to 'leadership for inclusion'.			

b. Teachi

This is the preview of the survey. No data entered into survey during preview will be saved.

	Yes/No	If yes, how extensive?	If yes, please briefly describe:
Provision to close the gap Improving the quality of teaching, assessment and target setting in order to improve and accelerate progress for all pupils.			
Conducting effective progress meetings with students Making the best use of time, and/or developing teachers' professional dialogue as a means to enhance learning outcomes for all pupils.			
Effective feedback Improving the quality of marking, assessment and feedback procedures in order to improve the learning, progress and wider outcomes of all pupils.			
Identifying and supporting SLCN Understanding the prevalence and range of speech, language and communication needs so that they can identify and support pupils to positively impact on their engagement and achievement.			
Literacy and English Developing effective whole-school approaches to reading, writing, speaking, and listening skills for all pupils.			
Numeracy and Maths Developing effective whole-school approaches to numeracy and maths skills for all pupils.			
Digital literacy Exploring the use and application of digital technology in educational settings.			
Other Please complete, if you had any other CPD pertaining to 'teaching and learning'.			

c. Wider

This is the preview of the survey. No data entered into survey during preview will be saved.

	Yes/No	If yes, how extensive?	If yes, please briefly describe:
Developing behaviours for learning, attendance and wellbeing Involving review of inclusive settings and strategies to support the participation, enjoyment and achievement of pupils in all areas of school life.			
Anti-bullying Review and development of effective anti-bullying strategies in school.			
Community engagement and partnership working Exploring and developing community and partnership working in education settings to improve provision and outcomes for all pupils, but particularly the vulnerable and disadvantaged.			
Developing pupil self-esteem Developing strategies to enhance your pupils' views of themselves.			
Developing resilience and self-efficacy Developing strategies to improve your pupils' ability to 'bounce back' in situations of adversity or challenge, and their belief that they can succeed in school and life.			
Managing transitions Exploring creative practices that effectively support pupils and families during major changes in their lives, as a means to improve their confidence, aspirations and achievements.			
Other Please complete, if you had any other CPD pertaining to 'wider outcomes and opportunities'.			

d. Paren

This is the preview of the survey. No data entered into survey during preview will be saved.

	Yes/No	If yes, how extensive?	If yes, please briefly describe:
Communication with parents Examining approaches to active communication/engagement with pupils' parents or carers to improve their involvement in their child's learning and achievement.			
Welcoming and including families Developing a culture and environment that is welcoming to, and inclusive of, all parents and carers.			
Other Please complete, if you had any other CPD pertaining to 'parent and carer engagement'.			

2. Please indicate the proportion of CPD noted above that has been provided by the following (please note that the total responses has to equal to 100%):

Higher Education Institution	
Local Authority	
Private consultant(s)	
Within school (e.g. SENCO)	
AfA 3As	
Other (please specify below*)	
Total	

*If other, please describe:

3. What

This is the preview of the survey. No data entered into survey during preview will be saved.

	HEI	LA	Consultants	Within school (e.g. SENCO)	AfA	Other
Workshops/lectures (e.g. single/series of workshops, conferences)	☐	☐	☐	☐	☐	☐
Coaching (e.g. mentoring, critical friendships)	☐	☐	☐	☐	☐	☐
Demonstrations/observations (e.g. demonstration lesson/videos, classroom observations, job shadowing)	☐	☐	☐	☐	☐	☐
Accredited training programme (e.g. Higher Education programmes, school-university partnership)	☐	☐	☐	☐	☐	☐
Networking (e.g. learning networks/collaboration with other schools, informal networking)	☐	☐	☐	☐	☐	☐
Within school training (e.g. INSET days, secondment/sabbaticals,)	☐	☐	☐	☐	☐	☐
Practitioner research projects	☐	☐	☐	☐	☐	☐
Online learning	☐	☐	☐	☐	☐	☐

Appendix L: UPS data baseline comparison

Table A39: Baseline comparison of school UPS measures

	Intervention			Control			
	N	Mean	SD	N	Mean	SD	Effect size
Total UPS score	66	9.80	3.20	68	9.41	3.26	0.12
Leadership for inclusion	66	12.88	6.91	68	12.10	7.67	0.11
Teaching and learning	66	19.47	9.04	68	17.60	8.38	0.21
Wider outcomes and opportunities	66	12.79	8.17	68	11.62	8.35	0.14
Parent and carer engagement	66	3.89	4.00	68	3.32	3.53	0.15

Note. The total UPS score is derived using 14 items from the survey (those suggested in the EFA)

Appendix M: UPS sub-category descriptives and ANOVAs

Table A40: Descriptive statistics for the two trial groups at DP1 and DP2

	Intervention		Control	
	DP1 Mean (SD)	DP2 Mean (SD)	DP1 Mean (SD)	DP2 Mean (SD)
UPS total score (14 items)	9.84 (3.23)	9.33 (3.99)	9.75 (2.91)	10.71 (3.52)
Leadership for inclusion	13.22 (7.08)	16.29 (9.66)	11.96 (7.24)	17.93 (9.73)
Teaching and learning	19.42 (9.43)	16.95 (10.34)	17.87 (8.37)	20.84 (9.47)
Wider outcomes and opportunities	12.85 (8.22)	12.36 (8.78)	11.91 (8.05)	15.16 (8.43)
Parent and carer engagement	4.13 (4.14)	5.07 (3.80)	3.51 (3.61)	5.38 (4.08)

UPS total score used in MLMs (using the 14 items suggested from the EFA):

- No main effect of time ($F(1, 108) = .296$, $p = .588$)
- No main effect of trial group ($F(1, 108) = 1.637$, $p = .203$)
- No interaction between group and time ($F(1, 108) = 3.103$, $p = .081$)

Leadership for inclusion:

- **Main effect of time ($F(1, 108) = 20.207$, $p < .001$)**
- No main effect of trial group ($F(1, 108) = .022$, $p = .881$)
- No interaction between group and time ($F(1, 108) = 2.068$, $p = .153$)

Teaching and learning:

- No main effect of time ($F(1, 108) = .049$, $p = .826$)
- No main effect of trial group ($F(1, 108) = .690$, $p = .408$)
- **Interaction between group and time ($F(1, 108) = 5.964$, $p = .016$)**

Wider outcomes and opportunities:

- No main effect of time ($F(1, 108) = 1.856$, $p = .176$)
- No main effect of trial group ($F(1, 108) = .566$, $p = .454$)
- No interaction between group and time ($F(1, 108) = 3.408$, $p = .068$)

Parent and carer engagement:

- **Main effect of time ($F(1, 108) = 10.102$, $p = .002$)**
- No main effect of trial group ($F(1, 108) = .066$, $p = .797$)
- No interaction between group and time ($F(1, 108) = 1.094$, $p = .298$)

Appendix N: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found **here**. Cost ratings are awarded as follows:

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix O: Security classification of trial findings

OUTCOME: Year 5 reading, whole group

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%	5				5
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **[0]**		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	No significant risk apparent.
Threat 2: Concurrent interventions	Low	The intervention was difficult to define, and some aspects may have been going on concurrently. However, this does not appear to have threatened the validity of the result
Threat 3: Experimental effects	Low	No significant risk apparent.
Threat 4: Implementation fidelity	Moderate	The intervention fidelity might have been lower than ideal, but since this is an effectiveness study, this does not seem to be a threat to validity.
Threat 5: Missing Data	Low	Low levels of missing data, and no significant effect.
Threat 6: Measurement of outcomes	Low	KS2 reading scores likely to be reasonably robust measures.
Threat 7: Selective reporting	Low	No evidence of selective reporting.

- **Initial padlock score:** 5 Padlocks
- **Reason for adjustment for threats to validity**: no threats to validity
- **Final padlock score:** 5 Padlocks

OUTCOME: Year 5 reading, target group

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%	5				5
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **[0]**		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Moderate	There was an imbalance at baseline of 0.1, which was controlled for in the primary analysis, with sensitivity analyses replicating the result.
Threat 2: Concurrent interventions	Low	The intervention was difficult to define, and some aspects may have been going on concurrently. However, this does not appear to have threatened the validity of the result
Threat 3: Experimental effects	Low	No significant risk apparent.
Threat 4: Implementation fidelity	Moderate	The intervention fidelity might have been lower than ideal, but since this is an effectiveness study, this does not seem to be a threat to validity.
Threat 5: Missing Data	Low	Low levels of missing data, and no significant effect.
Threat 6: Measurement of outcomes	Low	KS2 reading scores likely to be reasonably robust measures.
Threat 7: Selective reporting	Low	No evidence of selective reporting.

- **Initial padlock score:** 5 Padlocks
- **Reason for adjustment for threats to validity**: no significant threats to validity (two moderate risks of bias in unclear/ inconsistent direction)
- **Final padlock score:** 5 Padlocks

This document is available for download at **https://educationendowmentfoundation.org.uk**

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn