Education Endowment Foundation

Achieve Together
Evaluation report and executive summary
July 2017

Independent evaluators:

Ellen Greaves, Barbara Sianesi, Luke Sibieta, Neil Amin-Smith, Meg Callanan and Ruth Hudson

Institute for Fiscal Studies

NatCen Social Research

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

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

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;

- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and

- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) 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:

Danielle Mason
Head of Research and Publications
Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

p: 020 7802 1679
e: danielle.mason@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

The project was independently evaluated by a team from the Institute for Fiscal Studies (IFS): Ellen Greaves, Barbara Sianesi, Luke Sibieta and Neil Amin-Smith; and the process evaluation was led by NatCen Social Research: Meg Callanan and Ruth Hudson.

The lead evaluator was Ellen Greaves.

Contact details:

Institute for Fiscal Studies
3rd Floor
7 Ridgmount Street,
London,
WC1E 7AE

Email: mailbox@ifs.org.uk
Tel: 020 7291 4800

Contents

Executive summary

The project

Achieve Together aims to improve pupil outcomes by supporting leadership development and collaboration within schools in disadvantaged areas. The programme was devised and delivered by three education charities, Teach First, Teaching Leaders, and The Future Leaders Trust. It offers subsidised leadership development training using programmes provided by Teaching Leaders and The Future Leaders Trust, and places graduates into schools via the Teach First initiative. Participants on these individual programmes also collaborate on a school improvement initiative, drawing on insights from the individual programmes with planning and implementation support from the Achieve Together team. It is a whole-school intervention aiming to improve outcomes for all pupils in participating schools.

Achieve Together was developed in 2013 and delivered from September 2014 until July 2016. This efficacy trial assessed the impact on GCSE outcomes at the end of the second year of the programme. Results from 14 schools that received the intervention were compared to those of a well-matched comparison group. Qualitative research aimed to track implementation, delivery, and perceived impacts over the course of the intervention.

Key conclusions
1. Pupils in Achieve Together schools made no additional progress, on average, compared to other children in the trial, as measured by GCSE results. The three padlock security rating means we are moderately confident in this result.
2. Similar results were found for children eligible, and not eligible, for free school meals, and children with higher and lower prior attainment, although these sub-group results are less secure because of the smaller numbers of pupils.
3. Participating schools reported a lack of clarity over the aims of Achieve Together and how the collaborative project should relate to the individual leadership programmes. The differing timelines of the individual programmes also appears to have hampered the collaborative projects.
4. Views on the collaborative project were mixed. Some participants found that the process increased reflective practice and provided impetus for further school improvement. Others felt that it was resource intensive and had minimal additional impact.
5. Some teachers felt Achieve Together would have a positive impact on pupil attainment once new teaching skills and new initiatives were embedded. The EEF data archive allows analysis of future results for these schools to assess whether this is the case.

Security rating: 🔒🔒🔒🔒🔒

EEF security rating

These findings have moderate security. The security rating indicates how confident we can be that any additional progress experienced by the children in Achieve Together (AT) schools was due to the intervention and not any other factors.

The trial was an efficacy trial that tested whether the intervention can work under ideal or developer-led conditions. It was initially designed as a randomised controlled trial, but low recruitment led to a change in design to a well-matched comparison group study in 14 schools. The trial was well-powered, and relatively few pupils who started the trial were not included the analysis, for example, because they moved school.

A moderate security rating was awarded because the group of schools that signed up for the intervention were compared with a group of schools with similar characteristics, but which did not sign up. This means that although the two groups of schools look similar, there is a risk that there is an underlying systematic difference between them, as demonstrated by their willingness (or not) to sign up for this type of intervention. This reduces our confidence that any additional progress made by pupils in the AT schools is due to the intervention rather than to other factors such as underlying differences between the groups. To reduce the likelihood of such differences, comparison schools were selected from those that expressed an interest even though they did not actually sign up. In addition, the analysis shows that the pupils in the two groups were similar (balanced) with respect to the pupil characteristics that were measured. However, the risk of systematic differences remains.

Additional findings

There is no evidence that Achieve Together had an impact on pupil attainment at the end of secondary school, as measured by GCSEs and all secondary outcomes. This is true for all sub-groups of pupils considered.

Findings from the qualitative interviews were mixed. School staff believed that the leadership skills developed from participating in individual leadership programmes would ultimately feed into improved pupil attainment. However, it was felt that this would take time and that improvements would not be apparent immediately. In instances where schools had successfully implemented collaborative impact initiatives, the programme was felt to give added impetus to the school and increase reflective practice. Some schools found it difficult to achieve the alignment of individual participant activity that was expected as part of the wider school improvement collaboration, and some staff felt that the collaborative element had minimal impact. The process evaluation identified the following ways in which Achieve Together could be developed in the future: greater clarity of aims and expectations; sufficient planning and lead-in time; review of lengths and structures of individual programmes to improve compatibility; and more opportunities for collaboration across schools.

Cost

As a whole-school initiative, the per pupil cost of Achieve Together is low at £16.47 per year per pupil over three years, on average. There was variation in cost across the schools depending on the number of participants for each element of the programme. The range of total cost over three years per school is £17,000 to £75,000. The total average staff time required for a school participating in Achieve Together over three years was 84 days in school time and 188 days outside school time. This included the time taken for participating in the individual programmes and varied according to the intensity of the programmes.

Executive Summary Table: Summary of impact on primary outcome

Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	P value	EEF cost rating
GCSE in Achieve Together schools vs. Control	0.01 (-0.06; 0.09)	0	🔒🔒🔒⬜⬜	0.72	**£** £ £ £ £
GCSE for FSM vs. Control	0.00 (-0.08; 0.08)	0	-	0.93	**£** £ £ £ £

Introduction

Intervention

Achieve Together is an initiative devised and delivered by three education charities—Teach First, Teaching Leaders, and The Future Leaders Trust—to support leadership development and collaboration within schools to foster whole-school improvement and ultimately improve outcomes for children. The initiative aims to do this through two key means. First, Achieve Together aims to increase investment in training and teacher development by subsidising leadership development training for teachers in middle and senior leadership roles and by placing graduates into schools via the Teach First initiative. Second, Achieve Together offered a range of support to facilitate collaboration and alignment across these programmes, with a specific focus on participants working collaboratively on a school improvement impact initiative.

Achieve Together was devised and run by the AT team, which comprised representatives from each of the three constituent charities. Three different versions of Achieve Together were tested:

- an in-school initiative aimed at supporting leadership development and collaboration within an individual school;
- a whole-area programme in Bournemouth incorporating collaboration across schools in the area; this was funded by the EEF and J. P. Morgan, and will be reported on separately;[1] and
- two whole-area programmes focused on school-to-school collaboration in the North East and North West of England, funded by the Esmée Fairbairn Foundation.

In this evaluation, we report on the impact of the in-school initiative. In so doing, our evaluation will provide evidence on whether increasing investment in the training and development of teachers, when combined with activities to increase the strategic alignment of the different training programmes, is effective.

Achieve Together in-school initiative

Over the two-year period from September 2014 to July 2016, each participating school was expected to work with all three partner charities (Teach First, Teaching Leaders and The Future Leaders Trust). Each school was also encouraged to nominate between 8 and 12 participants for training courses run by these partner charities. Schools had autonomy in the choice and number of staff.

The intervention was focused on schools in low-income communities across England that were not already working with all three partner charities (see 'participant selection' for further details). London schools were excluded because of their significant existing participation with these charities. At the start of the programme, over 50% of schools eligible for the programme were rated as 'Requires Improvement' or 'Inadequate' by Ofsted.

The focus of the Achieve Together in-school initiative delivered by the three partner charities was:

- Leadership Development Training Programmes - schools were offered National Professional Qualifications and bespoke leadership development training for teachers at different stages of their teaching career, at subsidised rates; and
- Strategic Alignment Activities —the Achieve Together team offered a range of support and facilitation to increase strategic alignment between leadership development training and whole-school improvement priorities.

---

[1] https://educationendowmentfoundation.org.uk/our-work/projects/achieve-together-bournemouth/

In what follows, we describe each of these two main aspects of the programme in more detail.

Leadership Development Training Programmes

The core element of Achieve Together is the provision of five leadership development training programmes working concurrently with teachers in the same school. The content of the leadership development programmes was designed and delivered by the three partner organisations. They were not modified for the purposes of Achieve Together and were therefore the same as delivered to schools that participated in each programme separately. The three charities delivered the following programmes:

- Teach First places carefully selected trainee teachers in schools, usually to teach high-demand subjects—such as Physics, Maths, and Modern Foreign Languages—where schools find it difficult to fill vacancies. These trainees work towards qualified teacher status (QTS) and a PGCE qualification while teaching and being paid a salary.
- Teaching Leaders provides specialist training to promising middle leaders that includes coaching and support, formal training outside school hours, and an 'impact project' designed to improve grades and pupil life chances within the middle leader's area of responsibility. Teaching Leaders delivered two programmes as part of Achieve Together:
    - National Professional Qualification for Middle Leadership (NPQML), a nationally recognised qualification for middle leadership; and
    - Teaching Leaders Fellows, a programme designed for 'high-potential' middle leaders.
- The Future Leaders Trust develops existing senior members of staff through training programmes that support aspiring headteachers and the practice of senior leaders, or places senior leaders in schools. It delivered two programmes as part of Achieve Together:
    - National Professional Qualifications for Senior Leadership (NPQSL), a nationally recognised qualification for senior leadership within a school; and
    - Future Leaders programme, aiming to give participants the skills and strategies to lead schools in challenging circumstances.

Schools involved in the Achieve Together intervention signed a memorandum of understanding committing them to purchasing at least one programme from each of the three main charities. To further encourage schools to participate in these programmes, they were offered subsidised rates.[2]

It is important to note that schools could already be involved with one of the three charities to be eligible for the programme, however, if they were involved with more than one, they were not eligible. In Table 1 below, we show the proportion of treatment schools that were involved with each of these charities before the intervention, as well as the proportion of schools in our matched control group (defined in the methodology section below). As can be seen, participation with Teach First was relatively common among both groups, and clearly higher among the treatment schools. None of the treatment schools was involved with the other two programmes, while the matched control group had some limited involvement.

Achieve Together, therefore, aims to increase the provision of new high-potential teachers as demand requires, and train and develop existing staff. Given that most schools were not involved with Teaching Leaders or The Future Leaders Trust before the intervention, this is likely to comprise the biggest shift in training.

---

[2] Schools received a subsidy of £250 per participant from the EEF towards the cost of the Teaching Leaders (NPML) programme, and £2,000 per participant towards the cost of the Teaching Leaders (Fellows) programme. The EEF also provided a subsidy of £500 per participant for Future Leaders NPQSL. Neither Teach First nor the separate Future Leaders programme were subsidised by the EEF.

Table 1: Participation of Achieve Together schools and matched control group schools in individual programmes prior to the trial

Item	Achieve Together schools	Matched control group
Teach First	71.4%	47.5%
Teaching Leaders	0%	10%
Future Leaders Trust	0%	5%

Strategic Alignment

The second significant area of focus for the in-school intervention was a range of support and facilitation from the Achieve Together team to ensure that the leadership development training was well aligned to the existing priorities and context of the individual schools. The programme was developed in line with a 'Collective Impact' approach, prioritising continuous communication and co-creation of solutions in line with the needs of all stakeholders.[3]

Participants in the training programmes selected impact projects in areas of school improvement that related to their areas of responsibility. These formed part of the assessment requirements for the Future Leaders, Teaching Leader Fellows, NPQML, and NPQSL programmes. The main innovation of the Achieve Together programme was the intention to increase the strategic alignment between these impact projects.

A range of scalable activities were implemented by the Achieve Together team (see Appendix C) to increase alignment between the individual leadership development training and whole-school improvement priorities. Each activity was developed with the specific aim of supporting schools to lead whole-school projects themselves, with limited intervention from the three charities. In summary these activities were:

- a needs analysis process, involving an interview with the headteacher and analysis of key strategic documents to identify school improvement priorities; and
- a collaborative planning meeting facilitated by the Achieve Together project team and attended by all staff participating in Achieve Together leadership programmes; this included the use of the 'Business Canvas Model'—a tool adapted by Achieve Together to aid schools in planning their collective impact initiatives.
- In the second year of implementation, additional activities were developed by the Achieve Together team in response to the needs of schools (identified through the needs analysis process). This involved access to a range of CPD resources including:
  - the 'route map'—an online resource of CPD materials and elements of action research focusing on three areas of school improvement (formative assessment, coaching, and learning behaviours); and
  - a series of five webinars covering a range of school improvement topics (leadership and coaching, evidence-based approaches to self-improving schools, evidence-based practice and school improvement, formative assessment, and the Mental Toughness Questionnaire).[4]

More details of the individual activities can be found in Appendix C.

---

[3] For more details on the collective impact approach, see http://www.fsg.org/ideas-in-action/collective-impact
[4] The Mental Toughness Questionnaire is a diagnostic tool developed by AQR to assess mental toughness **http://aqrinternational.co.uk/mtq48-mental-toughness-questionnaire**. The tool was introduced to participating schools as a diagnostic tool for both staff (to measure leadership mind-sets) and pupils (to help identify ways to foster mental resilience and improve attainment).

Modifications to the trial

This trial was significantly modified over time. Originally, the trial aimed to test the impact of the Achieve Together in-school initiative using a randomised controlled trial involving 24 treatment and 24 control schools starting in September 2013. Although the initiative targeted improvements in pupil attainment across the whole school, we originally planned to focus on the attainment of Year 7 and 8 students as our primary outcome as new teachers are more likely to teach younger age groups. At this initial stage, GCSE results were intended as an additional outcome.

Initial recruitment was low and only 13 schools signed up to start the Achieve Together programme in September 2013 (see Recruitment section for further details). All these schools received the programme as planned. We refer to these schools as 'Cohort 1' but do not analyse the effect of AT on this group as no control group was established.

A second effort was made to establish a randomised controlled trial starting in September 2014, with slightly more generous subsidies for some of the programmes.[5] Again, insufficient numbers of schools could be recruited, with 14 schools willing to be part of a trial starting in September 2014. We refer to this group of schools as 'Cohort 2'.

Cohort 2 schools also received the programme as planned. However, the design of the evaluation was modified from a randomised controlled trial to a 'well-matched comparison group' (with the control group made up of schools thought to be similar to the intervention group, that is, those that expressed some interest in the programme but ultimately declined to participate). The primary outcome was changed from Year 7 and 8 pupil outcomes to GCSE outcomes (with the latter decided to keep the costs down). The change in the target group of pupils was not factored in prior to the recruitment of teachers onto leadership development programmes. While in principle the programme was expected to benefit all pupils, in practice we do not know the extent to which participants were directly involved in teaching GCSEs during the two years of the delivery.

The change in evaluation methodology led to a delay in engagement with schools. This meant that the first Achieve Together activities with schools took place from September 2014 rather than during the previous spring term. This change in the evaluation meant that collaborative planning took place after individual impact projects had been chosen by teachers whereas these were meant to be chosen after collaborative planning.

The trial is an efficacy trial, although the involvement of the Achieve Together team and design of the strategic alignment activities have been modified and developed throughout the programme in response to school needs and priorities

Background evidence

Teach First, Teaching Leaders, and Future Leaders are high-profile education charities working in a large number of schools in England. Up to 2014/2015, Teach First had partnered with 979 primary and secondary schools, and between them, Future Leaders and Teaching Leaders have worked in over 2,000 schools (these figures will include some Cohort 1 and Cohort 2 schools that were already participating with one or two of these charities). Each charity receives funding from the National College for Teaching and Leadership (NCTL) in addition to funding from sponsors. Each charity has separately commissioned or conducted evaluations of effectiveness, finding positive results. For example, 65.7% of disadvantaged pupils achieve expected progress in schools where headteachers are involved with

---

[5] This equated to an offer of broadly an *additional* £2,000 for each Teaching Leaders Fellow and £250 for each NPQSL participant. Control schools were also offered a £2,500 subsidy at the end of Year 2 of the intervention.

Future Leaders, compared to 58.2% in comparable schools (The Future Leaders Trust, 2015). For Teaching Leaders, 43.8% of pupils with low prior attainment make expected progress in maths where Teaching Leaders fellows are present, compared to 33.4% in similar eligible schools.[6] Using a matched comparison design, Allen and Allnut (2017) found that the introduction of Teach First leads to a school-wide increase in pupil attainment of around 5% of a standard deviation (or one grade in one of the pupil's best eight subjects). Comparing across departments with and without a Teach First participant also shows positive effects in the second year after involvement with Teach First, although not statistically significant positive effects for maths.[7] The evaluation by Allen and Allnut (2017) is more empirically robust than the evaluations of Teaching Leaders and The Future Leaders Trust. Allen and Allnut construct a well-matched comparison group from schools in the same region that take part in Teach First in later years and carefully account for pre-programme trends in attainment.[8]

The Achieve Together trial is the first to provide evidence on whether there is an accelerated impact on pupil attainment resulting from the three charities working in partnership in disadvantaged schools in England. The results therefore suggest whether an intensive human-capital approach to school improvement is effective.

Evaluation objectives

The objectives of the impact evaluation were to measure the impact of Achieve Together on attainment at Key Stage 4 and pupil absence using quasi-experimental methods to construct a well-matched comparison group for the Achieve Together schools. The objective of the process evaluation was to understand the implementation and delivery of the intervention over the first two years of the programme, including staff reflections on the programme, facilitators and barriers to engagement, impact and key learning, and recommendations for programme roll-out.

The process evaluation was structured around the key components of the initiative, drawing on the theory of change developed by the Achieve Together project team (Appendix D). This has been modified since originally planned due to the change in the focus of the trial.

The impact and process evaluations follow the original evaluation protocol for Achieve Together, available at:

https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Achieve_Together.pdf

Ethical review

No pupil consent procedures were necessary for the impact evaluation as there was no processing of sensitive personal data outside of the National Pupil Database, which has strict conditions for use and access. Schools did, however, sign a memorandum of understanding (Appendix E) where they committed to providing necessary information and data for both the impact and process evaluations. The process evaluation underwent an ethical review with NatCen's Research Ethics Committee and received approval in October 2014, prior to data collection. Even though the trial design changed, treatment schools were not asked to sign a different memorandum of understanding as no further information or commitment was required from schools.

---

[6] http://www.teachingleaders.org.uk/wp-content/uploads/2016/03/Secondary-Impact-2-pager-20161.pdf. The equivalent figures for English are 70.5% in departments with Teaching Leaders fellows, compared to 55.8% in eligible schools.
[7] http://repec.ioe.ac.uk/REPEc/pdf/qsswp1311.pdf; IFS CBA report
[8] See McIntyre and Thomson (2016) for a case study of Teach First.

Project team

Achieve Together was initially developed collaboratively by the three education charities, Teach First, Teaching Leaders, and Future Leaders. The development of, and support for, the Collective Impact Project was led by Chris Gurney and Zoya Wallington at Teach First.

The evaluation was directed by Ellen Greaves. The impact evaluation was led by Ellen Greaves, with Barbara Sianesi, Luke Sibieta and Neil Amin-Smith, all from the Institute for Fiscal Studies. The process evaluation was led by Meg Callanan with Ruth Hudson, both at NatCen.

Methods

Trial design

The most empirically robust way to test the impact of the Achieve Together programme would be through a Randomised Controlled Trial (RCT), and this was the original intention. The initial design was an RCT: the intention was to recruit at least 48 schools, 24 of which would be randomly assigned to the treatment group. The number of schools eventually recruited, however, prohibited an RCT design. Recruitment to Achieve Together was lower than anticipated for a number of reasons: the number of eligible schools (able to work with all three charities, not working with more than one, outside London) provided a small pool of schools; concerns about being allocated to the control group when interested in all three charities; concerns about capacity to support the delivery of the intervention; the relatively high cost of involvement; uncertainty about staff numbers to participate; uncertainty about the school budget; and the complexity of the offer.

In its place, we sought to find the most empirically robust non-experimental way to assess the impact of Achieve Together. In particular, we sought to create a well-matched comparison group for the 14 schools in the intervention group that received the Achieve Together programme between September 2014 and July 2016 (Cohort 2 schools). The well-matched comparison group is taken from the group of 72 schools that were eligible for Achieve Together (on the basis of pupil disadvantage) and expressed some interest in Achieve Together before deciding not to participate, either expressing concern about participating in the trial (as was originally proposed), or due to financial considerations. These schools are 19% of the schools that were eligible for Achieve Together but did not sign up (13% from financial considerations and 6% from concern about participation in the trial). This group of schools is likely to have similar characteristics to the intervention group as the Achieve Together programme was of interest to both groups of schools. In the results section, we confirm that observable characteristics and baselines outcomes are generally similar across intervention schools and those that merely expressed an interest.

There is, however, clearly potential for there to be *unobservable* differences across the two groups that could drive differences in outcomes after the intervention was delivered (reducing our ability to ascribe any differences between the two groups to be the result of Achieve Together). For example, schools most concerned about participation with the trial might be those that require intervention the most, and therefore have lower pupil outcomes, on average. Similarly, those with financial problems may be expected to have lower pupil outcomes, on average. Alternatively, these schools may have lower pupil outcomes in prior years but be expected to experience higher growth in future years.

With such concerns in mind, we also used propensity score matching to find a smaller set of schools (41) from those that expressed an interest in Achieve Together that have the most similar characteristics before any involvement with the intervention. Using this approach for the set of schools that expressed an interest in Achieve Together aims to ensure that schools in the intervention and comparison group are similar in observable characteristics (such as pupil composition and attainment) and unobservable characteristics (such as senior leaders' orientation to professional development and school improvement). Any differences between the outcomes for the intervention and control group can therefore be more reasonably interpreted as the impact of Achieve Together. However, some caution is still needed as similarity across unobservable factors is not guaranteed by this approach and cannot be tested.

The precise propensity score matching specification for selecting the 41 schools was decided in May 2016 (before final outcome administrative data was available) using baseline characteristics of schools (from the 2012/2013 academic year). The optimal matching specification was the specification with the lowest absolute percent median bias across key school variables from the 2012/2013 academic year. The matching specifications tested varied the independent covariates and propensity score matching

method. The final specification selected uses nearest neighbour matching with six nearest neighbours and caliper of one. The independent matching variables are as follows:

- continuous variable for percentage achieving five A*–C GCSEs and equivalents, including English and maths, in academic year 2012/2013 from the publicly available performance profiles (ptac5em);
- continuous variable for total average (capped) point score per pupil (GCSEs only) in academic year 2012/2013 from the publicly available performance profiles (ttapsgcp);
- continuous variable for value added measure based on the best eight GCSE and equivalent results in academic year 2012/2013 from the publicly available performance profiles (b8vamea);
- continuous variable for percentage of pupils with English as an additional language ('EAL pupils') taken from the publicly available school-level census in academic year 2012/2013 (pnumeal); and
- binary variable for latest Ofsted rating for 'leadership and management' at least 'good' in academic year 2012/2013.

The full set of models tested is given in the Statistical Analysis Plan (published alongside this report on the EEF website). This was decided without consideration of the primary and secondary outcome variables, indeed before the primary outcome variable was available. Pupil attainment is measured using the National Pupil Database (NPD) with measurements before the intervention and after the first and second year of the intervention. It will be possible to measure the longer-term impact of the intervention in later years as the well-matched comparison group will not receive Achieve Together at any point.

Schools participating in Achieve Together could already be working with one or two of the three partner charities, but not all three. It is possible that the treatment effect could vary with existing engagement, but unfortunately this is not possible to explore given the small sample size and availability of information. By virtue of the fact that control schools were selected after the intervention phase had finished, no restrictions could be placed on these and they could have unilaterally chosen to partner with the three partner charities. However, they would not have had access to the financial subsidies for training programmes, nor the strategic alignment aspects. All pupils in participating schools are considered to be treated by the intervention, although outcomes are measured only for pupils in the final year of secondary school where administrative data is available.

Participant selection

The population of intervention schools is the 14 schools across England that signed up to the Achieve Together trial (beginning in September 2014) with full knowledge of the original RCT design.

Participating schools were selected from schools that met the eligibility criteria used by Teach First, Teaching Leaders, and Future Leaders at the time: at least 50% of pupils needed to come from the lowest 30% of the IDACI (Income Deprivation Affecting Children Index), or at least 50% of pupils needed to be eligible for free school meals ('FSM pupils'), or have a performance at KS4 below the national average for schools where the proportion of FSM pupils is between 25% and 50%. An additional eligibility criterion was that the school could not currently be working with all three partner charities (but could have been working with one or two). Schools in London were excluded from recruitment given the existing high engagement with either Teach First, Teaching Leaders or Future Leaders.

Recruitment was led by Teach First. Contact was made with school representatives, and meetings were held to explain the aim of Achieve Together and the conditions of the evaluation (at the time an RCT). A school is considered an intervention school if they signed the Memorandum of Understanding (reproduced in Appendix E).

As already indicated, recruitment was not as high as expected across both cohorts. The main reasons schools gave for not wanting to participate were concerns regarding the cost of the programmes and

potentially being in the control group (and thus unable to access some of the programmes independently).

Outcomes measures

All primary and secondary outcomes are obtained from the NPD for all intervention schools and schools in the well-matched comparison group.

The primary outcome for the school-level intervention is capped GCSE and equivalents new style point score for KS4 students in Achieve Together schools after two years of the intervention, that is, those that started Year 10 in 2014. This variable is standardised using the whole population of KS4 pupils to have a mean of zero and standard deviation of one. The primary outcome is different from that specified in the original protocol which focused on Year 7 and 8 outcomes. As part of the shift from an RCT to a well-matched comparison group, a decision was made not to collect the Year 7 and 8 outcome data for reasons of cost, and to only use GCSE outcomes as the primary outcome (originally intended as a secondary outcome).

Secondary outcomes are:

1. Capped GCSE point score excluding equivalents. This outcome is standardised using the whole population of KS4 pupils to have a mean of zero and standard deviation of one. This outcome is included as an alternative measure of attainment at the end of secondary school that is not affected by a school's decisions around entry to alternative assessments.
2. Teach First's measure used for their second 'impact goal'.[9] This is the capped GCSE point score excluding equivalents plus English and maths bonuses. This outcome is standardised using the whole population of KS4 pupils to have a mean of zero and standard deviation of one. This outcome is included to be comparable with Teach First's internal evaluation.
3. Overall absence for KS4 pupils across the academic year. This outcome is included as a proxy for pupils' engagement in school.
4. Primary and secondary outcomes described above from Year 1 of the intervention (pupils who started Year 11 in September 2014 and took GCSE exams in summer 2015). These are included in the analysis as indicative of the change in attainment over the two-year intervention.

Sample size

Initial sample size calculations were completed for the original design of the intervention (randomised controlled trial with 48 schools). These sample size calculations suggested that with a realistic level of intracluster correlation of pupil outcomes within schools (0.2), the minimum detectable effect size (MDES) would be around 0.28. This assumed 170 pupils per year-group in 24 intervention and 24 comparison schools, with desired statistical power of 80% and significance level of 5%, and that the unexplained variance in the outcome variable is 0.6. The equivalent effect size for FSM pupils was 0.29, assuming around 45% of pupils were eligible for FSM in the past six years in intervention and matched comparison schools (based on previous years of NPD data).

This MDES is large relative to existing estimates. The sample size calculations gave a similar MDES once assuming a smaller intervention group and larger comparison group. For the same assumptions of pupils per year per school, statistical power, significance level, and unexplained variation in the outcome variable, but with 15 intervention schools and 30 comparison schools, the MDES was 0.31. For FSM pupils this was 0.32.

The final sample size was determined by the number of schools that signed up to Achieve Together.

---

[9] https://www.teachfirst.org.uk/why-we-exist/what-were-calling

Analysis

The analysis model is at the pupil level, comparing the outcomes for pupils in intervention and well-matched comparison schools. The comparison group was chosen from the schools that expressed an interest in being in the Achieve Together trial (discussed in the Trial Design section).

An additional analysis (specified in the Statistical Analysis Plan) tests the robustness of the estimates using the well-matched comparison group from the set of all schools that expressed an interest in Achieve Together. This analysis conducts the equivalent exercise for Achieve Together schools in Cohort 1 schools and compares it to an alternative comparison group likely to be even more similar in unobservable characteristics: the second cohort of AT schools as a comparison group to the first cohort. This is to inform whether the well-matched comparison group method using the secondary cohort of Achieve Together schools is likely to be unbiased. This analysis is presented in Appendix F.

As this is a well-matched comparison group design rather than an RCT, we use two further ways to ensure the comparison group and intervention schools are as similar as possible. First, we use probability weights derived from the optimal propensity score matching specification to give the best balance between intervention and well-matched comparison group.[10] Second, given that there remain differences in observable characteristics between intervention and comparison schools, we also control for pupil- and school-level covariates in the model. These are outlined below. The model was run in Stata 14.

Pupil-level covariates:

- binary variables for season of birth (Winter, Spring, Summer, with Autumn as reference category) from 2016 KS4 NPD data;
- binary variable for male from 2016 KS4 NPD data;
- binary variables for Ever FSM 6 ('yes' and 'missing', with 'no' as reference category) from 2016 KS4 NPD data;
- continuous variable for IDACI score and binary variable for missing IDACI score from 2016 KS4 NPD data—IDACI score has a large and statistically significant effect over and above FSM in previous years of NPD data; and
- binary variables for attainment at KS2 (lowest quartile, second lowest quartile, second highest quartile and missing, with highest quartile as reference category) from 2016 KS4 NPD data linked to prior attainment at KS2.

School-level covariates:

- continuous variable for percentage achieving five A*–C GCSEs and equivalents, including English and maths, in academic year 2012/2013 from the publicly available performance profiles (ptac5em);
- continuous variable for total average (capped) point score per pupil (GCSEs only) in academic year 2012/2013 from the publicly available performance profiles (ttapsgcp);
- continuous variable for value added measure based on the best eight GCSE and equivalent results in academic year 2012/2013 from the publicly available performance profiles (b8vamea)—value added, total average (capped) points score, and percentage achieving five A*–Cs have a relatively low correlation for the group of intervention schools which justifies inclusion of all three of these variables, between 0.34 and 0.88;
- continuous variable for percentage of teachers with Qualified Teacher Status in academic year 2012/2013 from publicly available school-level data derived from the School Workforce Census (qts) (the theoretical reasoning for this variable is that schools are likely to have signed up to Achieve Together to address staff shortages, which are proxied by this variable);
- continuous variable for average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching' in academic year 2012/2013;

---

[10] Common support is also imposed.

- binary variable for latest Ofsted rating for 'leadership and management' at least 'good' in academic year 2012/2013, as Achieve Together schools have particularly high scores on this scale;
- continuous variable for percentage of FSM pupils taken from the publicly available school-level census in academic year 2012/2013 (pnumfsm);
- continuous variable for percentage of EAL pupils taken from the publicly available school-level census in academic year 2012/2013 (pnumeal); and
- continuous variable for percentage of pupils with SEN statement or on School Action Plus taken from the publicly available school-level census in academic year 2012/2013 (psensap).

The effect size is calculated through converting each primary and secondary outcome variable into standard deviation units. The difference between the intervention and comparison group therefore gives the difference in outcome in standardised units.

Clustering at the school level (the correlation of pupil outcomes within schools) is taken into account using robust standard errors. A random effects/multi-level model is not used as the main model given the required assumption that the school-level effect is uncorrelated with all pupil- and school-level covariates in the model. This is not reasonable in this circumstance as pupil characteristics are likely to affect school effectiveness. A sensitivity analysis which includes random effects is included in Appendix F, however.

Sub-group analyses on the primary outcome are carried out on the following groups as specified in the evaluation protocol: whether or not a pupil has ever received free school meals (everFSM), and pupils with low prior attainment. Low prior attainment is defined as not achieving at least level 4 at Key Stage 2 in each of maths, English, and (teacher assessed) science. This is done using a model identical to the primary outcome model but splitting the sample to include the subgroup of interest only.

Missing outcome data is not imputed. Pupils with missing values for at least one primary and secondary outcome (aside from absence) are excluded (1% of pupils in 2016). The absence variable has more missing values (around 1.3% of the sample), and so the common sample is not restricted to non-missing values of this variable. Sensitivity analysis that restricts the common sample to include non-missing values of all primary and secondary outcomes (around 2% of pupils in 2016) is presented in Appendix F.

Missing values of independent variables are accounted for by binary variables. These binary variables capture the differences in outcome variable for those with and without missing values, on average, which are likely to be driven by the non-random selection. For example, missing KS2 values are likely to reflect not being present in the state sector in England at the end of primary school. (Missing KS2 values are more likely for the KS4 cohort in 2015—after the first year of Achieve Together— due to the KS2 boycott in 2010. Missing values for this cohort due to school-level boycott are imputed using teacher assessments in English and maths. A binary variable equal to one if these imputed variables are used, and zero if not, is also included. Sensitivity analysis is conducted using these variables for all students.)

In addition, and as specified in the Statistical Analysis Plan, we conduct a small element of 'on treatment' analysis. On treatment will be defined as having at least one participant from the school on the senior leadership training programme (NPQSL) over the two-year programme. This was proposed by the Achieve Together team as being crucial to the success of the programme. The analysis model will be then identical but with the four schools that do not meet this criterion excluded.

Implementation and process evaluation

A longitudinal qualitative research design was used to track implementation, delivery, and perceived impacts over the course of the pilot. Five case-study schools were selected and depth interviews were

conducted with staff participating in the pilot in the spring and summer terms of the first year, and the summer term of the second year, to track developments over time.

The process evaluation does not evaluate the individual leadership programmes (Teach First, NPQML, NPQSL, TL Fellows, and Future Leaders flagship programme) that formed the foundation of the Achieve Together model as this was beyond the scope of the evaluation (see Allen, R. *et al.*, 2017, Mujis, D. *et al.*, 2010, and Future Leaders, 2010 for evaluations of individual programmes). Rather, the focus was on the added value of the Achieve Together model—specifically the collaboration and alignment across the leadership programmes.

In total, 29 interviews were carried out with staff in participating schools in addition to four interviews with the Achieve Together project manager. This was fewer than the 48 interviews originally intended. This was in part because recruitment was challenging and not all schools responded to requests to participate and in part because the number of staff participating in the programmes in the case-study schools was lower than anticipated at the outset. In light of lower levels of recruitment than planned, all participating schools were approached to contribute in the final phase of fieldwork in the summer 2016, leading to a further two schools providing feedback on the pilot. Table 2a a summarises the number of participants across the programmes in the five case-study schools, and Table 2b provides a summary of achieved interviews.

Table 2a: Total number of participants who started programmes in case-study schools

	Teach First	Teaching Leaders	Future Leaders	Total
School 1	7	3 (Fellows)	0	10
School 2	5	11 (NPQML) 2 (Fellows)	5 (NPQSL)	23
School 3	5	9 (NPQML)	1 (NPQSL)	15
School 4	6	1 (Fellow)	2 (NPQSL)	9
School 5	6	2 (NPQML) 1 (Fellows)	2 (NPQSL)	11

Table 2b: Achieved interviews

	Year 1 spring 2015	Year 1 summer 2015	Year 2 summer 2016	Total
AT project manager	2	1	1	4
School 1	School lead x 1	School lead x 1	School lead x 1	3
School 2	School lead x 1	School lead x 1 NPQSL (paired) x 1 NPQML (group) x 1 NPQML (paired) x 1 Teach First (paired) x 1	School lead x 1 NPQSL (paired) x 1 NPQML (paired) x 1 Teach First (paired) x 1	10
School 3	School lead x 1	Teach First x 1	0	2
School 4	School lead x 1	School lead x 1	School lead x 1 TL Fellows x 1	4
School 5	0	School lead x 1 NPQSL x 1 TL Fellows x 1 Teach First x 2	School lead x 1 TL Fellows x 1	7
School 6	0	0	School lead x 1 NPQML x 1	2
School 7	0	0	School lead x 1	1
Overall total	**6**	**14**	**13**	**33**

Seventeen interviews were conducted face to face and sixteen by telephone. Interviews lasted between thirty minutes and one hour and were audio recorded and transcribed. Transcripts were analysed using the Framework approach—a systematic and comprehensive approach to data management that facilitates within and between case analyses of qualitative data.

Costs

We collected cost information from the Achieve Together team which was in turn collected from the three partner charities. This included the cost per participant for each programme for schools not involved the intervention (for example, excluding the subsidy provided by the EEF). This was not part of our original plan as this was not part of the EEF's standard approach at the time.

In addition, a survey of intervention schools aimed to capture the costs to a school associated with Achieve Together not directly captured by the individual programme costs, for example additional resource costs and travel costs. The response rate to the survey was very low (four schools), however, and so these costs are not included in the overall calculations of costs.

The cost is calculated as the cost per pupil per year averaged over three years for a school delivering Achieve Together outside the conditions of the trial, in line with EEF guidance. For example, schools in the intervention in the trial received a subsidy from the EEF of £2,000 for the cost of each Teaching Leaders Fellows participant, and this subsidy is not included in the cost calculation. There is variation in total cost per school across schools in the intervention; this relates to the number of participants from each programme that were required.

In line with EEF guidance for whole-school interventions, we have reported the cost per pupil (calculated as the cost per school divided by the average number of pupils per school), and the average cost of the

intervention over three years. The average number of pupils per school (931) was calculated using publicly available information on school size for each intervention school.

In line with EEF guidance we also report staff time for involvement with the intervention separately from the costs of the intervention, based on information provided by the three delivery organisations.

Timeline

Table 1: Timeline

Date	Activity
September 2013–March 2014	Achieve Together team recruited schools.
May 2014	Decision made to use well-matched comparison design rather than randomised controlled trial design.
September 2014	Start of Year 1 of Achieve Together.
March–April 2015	Process evaluation: Wave 1 interviews.
May 2015	Applied for access to NPD for KS4 results for June 2014 (baseline for KS4 attainment at the school level), absences and prior KS2 attainment.
July 2015	Process evaluation: Wave 2 interviews.
September 2015	Year 2 of the trial begins.
February 2016	Applied for access to NPD for KS4 results for June 2015 (Year 1 of the intervention), absences and prior KS2 attainment.
May 2016	Selection of optimal propensity score weights and set of well-matched comparison schools. Analysis of KS4 attainment and absences after the first year of the pilot.
June 2016	Presentation of interim findings.
July 2016	Applied for access to NPD for KS4 results for June 2016 (Year 2 of the intervention), absences and prior KS2 attainment. Process evaluation: Wave 3 interviews.
December 2016	Process evaluation: Analysis of interview data.
January 2017	Analysis KS4 attainment and absences after the second year of the pilot.
February 2017	Draft final report.
February 2017	Final report.

Impact evaluation

Participants

Figure 1 shows the number of schools and pupils at each stage of the evaluation; 14 of the 411 schools that were approached for Achieve Together signed the Memorandum of Understanding and participated in the programme (starting in September 2014). Administrative data for all of these schools was observed to analyse the impact of the intervention on pupil outcomes. A minority of pupils (about 1%) were excluded from analysis due to missing primary and secondary outcomes.

The well-matched comparison control group was selected from the 72 schools that were eligible for Cohort 2 of Achieve Together and had expressed an initial interest in being part of the programme, but in the end did not sign up for either financial considerations or because of concerns they would be part of the control group. From this group of 72 schools, we selected a group of 41 schools that were as similar as possible to the 14 treatment schools on the basis of observable school characteristics (precise methods are described in the 'Methodology' section above). This group of 41 schools represents our well-matched comparison group.

The other 313 eligible schools that did not express an interest in joining the programme are not analysed as part of this evaluation.

Table 4 shows the number of participants in each of the Achieve Together programmes by school (where schools have been anonymised). The total number of participants varies significantly, from a minimum of five to a maximum of 23. The most widely used programmes were Teach First and the Teaching Leaders NPQML course. Future Leaders' NPQSL course had just under two participants per school (with two schools dominating use here) and the Teaching Leaders Fellows around one per school. Future Leaders flagship programme was the least used, being only used by one school across the intervention. This suggests that there was clearly a lot of heterogeneity in terms of the leadership training programmes schools used, as was intended.

We saw earlier that most intervention schools were already participating with Teach First (just over 70%), though the commitments as part of Achieve Together are likely to have increased the level of participation here. Indeed, a figure of 4.6 participants per school is well above the average across each region of England, where the average intensity varies from around 1 to 3 per school.[11] No schools in the treatment group were participating in Teaching Leaders or Future Leaders programmes before Achieve Together began. This comparison suggests that we should probably interpret the effect of the intervention as mostly including increased use of Teaching Leaders programmes and an increase in the number of Teach First participants per school (though most were already engaging in Teach First to some extent to begin with). There was increasing use of Future Leaders' programmes, but this was not as great and mostly concentrated in two schools.

---

[11] http://graduates.teachfirst.org.uk/leadership-development-programme/where-you-could-work
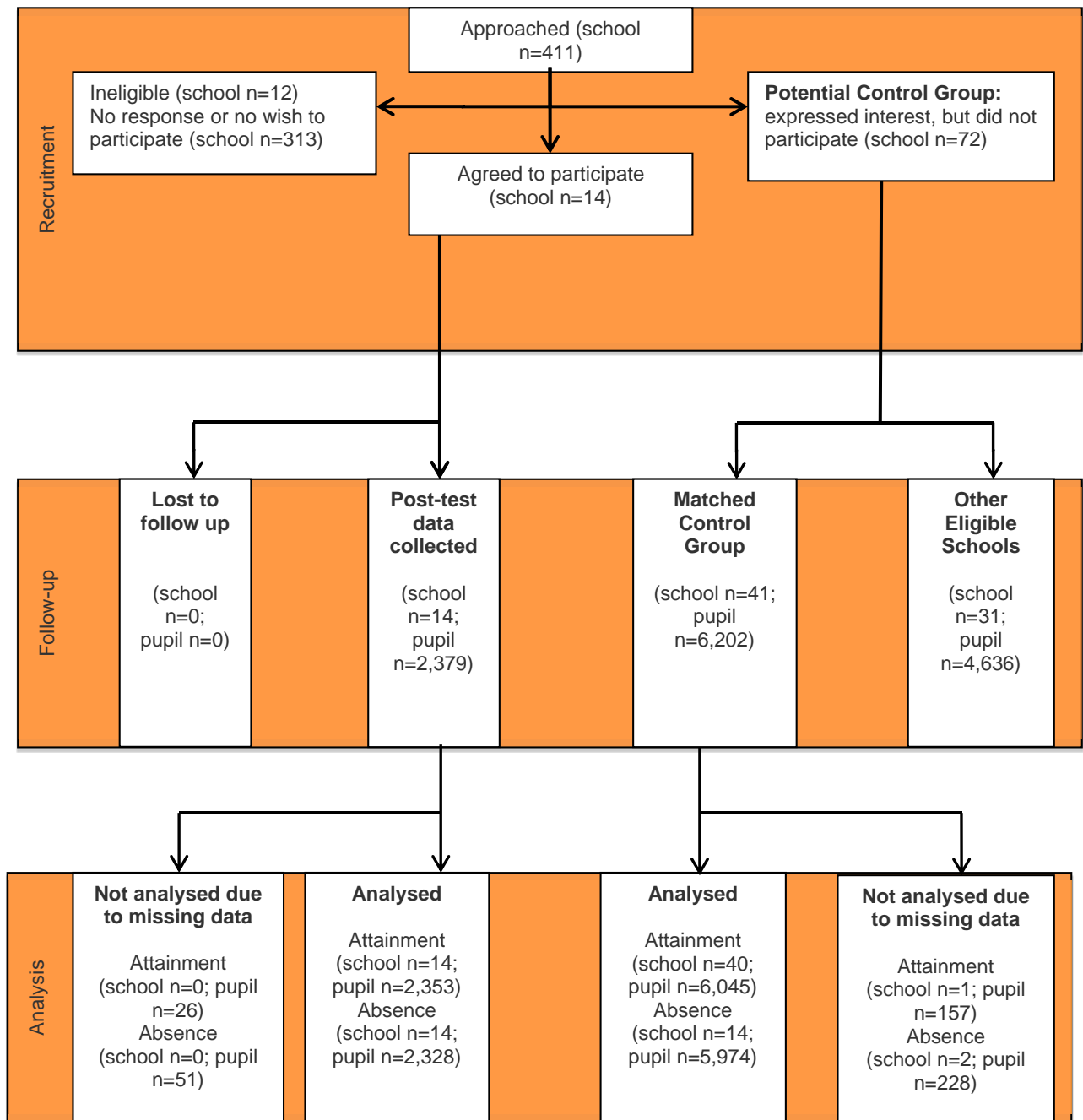
Figure 1: Participant flow diagram

Recruitment	Ineligible (school n=12) No response or no wish to participate (school n=313)	Approached (school n=411)		Potential Control Group: expressed interest, but did not participate (school n=72)
		Agreed to participate (school n=14)		

	Lost to follow up (school n=0; pupil n=0)	**Post-test data collected** (school n=14; pupil n=2,379)	**Matched Control Group** (school n=41; pupil n=6,202)	**Other Eligible Schools** (school n=31; pupil n=4,636)
Follow-up				

	Not analysed due to missing data Attainment (school n=0; pupil n=26) Absence (school n=0; pupil n=51)	**Analysed** Attainment (school n=14; pupil n=2,353) Absence (school n=14; pupil n=2,328)	**Analysed** Attainment (school n=40; pupil n=6,045) Absence (school n=14; pupil n=5,974)	**Not analysed due to missing data** Attainment (school n=1; pupil n=157) Absence (school n=2; pupil n=228)
Analysis				

Table 2: Number of participants in each programme by school (anonymised)

Item	Teach First	TL NPQML	TL Fellows	NPQSL	Future Leaders	Total Participants
School A	3	1	0	1	0	5
School B	7	0	3	0	0	10
School C	3	2	0	1	0	6
School D	6	4	0	7	0	17
School E	6	0	1	2	0	9
School F	6	2	1	2	0	11
School G	5	15	2	0	1	23
School H	6	10	0	0	0	16
School I	5	9	0	1	0	15
School J	7	0	0	0	0	7
School K	2	0	2	2	0	6
School L	5	11	2	5	0	23
School M	0	10	1	2	0	13
School N	4	0	0	1	0	5
Average	*4.6*	*4.6*	*0.9*	*1.7*	*0.1*	*11.9*

Note: Schools referred to in random order. Teach First and TL Fellows had new starters in both years of the programme and the number on these two programmes is the total starting across both years.

The minimum detectable effect size at different stages of the evaluation is presented in Table **3**. The evaluation protocol assumed 15 intervention and 15 comparison schools, with a correlation between pre-test and post-test of 0.6 and an intracluster correlation of 0.15. This resulted in a minimum detectable effect size of 0.312. This means that the effect of the intervention would have to be 0.312 standard deviations to be statistically significantly different from zero in a statistical test with a power of 80% and significance level of 5%.

Using the whole sample of schools that expressed an interest in Achieve Together as the comparison group, the number of schools is larger than specified in the evaluation protocol; this increases the power of statistical tests, all else equal. However, this is largely because of the change in the design of the trial, with the control group now comprising all schools that expressed an interest, but did not sign up. The intracluster correlation is also lower than expected (0.086) which increases the power of statistical test, all else equal. The resulting minimum detectable effect size is therefore lower, but still relatively large (0.189).

The analysis sample is smaller at 41 as we selected a smaller group of schools designed to be as similar as possible to the treatment group. As a result, the minimum detectable effect size for the analysis sample is slightly lower still, but remains large (0.182).

Table 3: Minimum detectable effect size at different stages

Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	30 / 5,100 (15 / 2,550; 15 / 2,550)	0.60	0.15	N/A	80%	0.05	0.312

Recruited	87 / 13,323 (14 / 2,379; 73 / 10,944)	0.58	0.086	N/A	80%	0.05	0.189
Analysis	55 / 8,650 (14 / 2,379; 41 / 6,271)	0.58	0.083	N/A	80%	0.05	0.182

Pupil characteristics

Table 5 shows the baseline demographic, attainment, school-level and other relevant characteristics of the intervention group and potential comparison group of schools (those that expressed an interest in Achieve Together). Baseline school-level variables are taken from publicly available school census data from 2013. Baseline pupil-level variables are taken from the NPD from the academic year 2013/2014. There are no statistically significant differences in school-level categorical or continuous variables between the intervention group and potential comparison group. This is encouraging, suggesting that the set of potential comparison schools is reasonably similar to the group of schools that signed up to the intervention. There are some relatively large differences, however, and the lack of statistical significance may be due to the small sample size of schools. For example, on average, 46% of pupils in intervention schools achieved five A*–C grades (including English and maths) in 2013 compared to 49% of pupils in potential comparison schools. The value-added score for interventions schools is noticeably lower, and the percentage of EAL pupils is almost 10% higher. For pupil-level baseline characteristics, the intervention and potential comparison group are well-balanced according to a measure of local deprivation (IDACI), but there are significant differences in the percentage of pupils currently and ever eligible for free school meals (which are both higher in intervention schools).

These comparisons represent differences in school characteristics at one point in time. It is possible, however, that outcomes across the groups were subject to differential trends before the trial and this could have affected schools' willingness to participate. If this were the case, such differential trends could confound our estimates of the impact of Achieve Together. For example, schools that signed up for the trial could have been those who had just experienced a dip in school performance and might thus be expected to bounce back during the trial. Appendix Figure 1 shows the trends in school performance for the six years prior to the trial. While it is clear that Achieve Together schools and the potential comparison group (interested and eligible schools) have slightly different levels of performance in any given year, they both appear to follow the same trend over time. This adds further weight to the idea that the potential comparison group was very similar to the group of Achieve Together schools prior to the start of the trial.

Table 4: Baseline comparison (unmatched sample)

Variable	Intervention group		Potential comparison group	
School-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Ofsted overall effectiveness outstanding or good	8/14 (0)	57.1	36/66 (6)	54.5
Ofsted quality of teaching outstanding or good	8/14 (0)	57.1	38/66 (6)	57.6
Ofsted leadership and management outstanding or good	9/14 (0)	64.3	47/66 (6)	71.2
School-level (continuous)	**N(missing)**	**Mean**	**N (missing)**	**Mean**
% 5A*–C (including English and maths)	14 (0)	46.1	72 (0)	49.2
Capped points score	14 (0)	221	72 (0)	218.7

Value added	14 (0)	1007.4	72 (0)	1001.8
Teachers with QTS (%)	14 (0)	93.2	70 (2)	94.6
% of pupils eligible for FSM	14 (0)	32	72 (0)	31.5
% of pupils with EAL	14 (0)	28.8	71 (1)	19.6
% of pupils with SEN	14 (0)	10.6	72 (0)	11.6
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Male	1,270/2,441 (0)	52.0	6,075/11,864 (0)	51.2
Ever received FSM*	1,151/2,441 (0)	47.2	5,358/11,864 (0)	45.2
Eligible for FSM***	740/2,441 (0)	30.3	3,147/11,864 (0)	26.5
Pupil-level (continuous)	**N(missing)**	**Mean**	**N (missing)**	**Mean**
IDACI score**	2,441 (0)	0.347	11,864 (0)	0.35
Note: Stars for statistical significance: p <= 0.10 (*), p <= 0.05 (**), p <= 0.01 (***).				

Table 6 shows the equivalent characteristics for the group of 41 well-matched comparison schools applying the optimal propensity score weights which are used throughout the analysis presented in the following tables (for further details on the selection of comparison schools, please refer to the Statistical Analysis Plan published alongside this report). For the majority of school- and pupil-level variables, the well-matched comparison schools are more similar than the total set of potential comparison schools. For example, the percentage of schools with overall effectiveness graded 'outstanding' or 'good' by Ofsted increases to 56.1% from 54.5%, which is closer to the 57.1% in the intervention group. There is no longer a significant difference in the percentage of pupils currently eligible for free school meals using the matched sample, and the difference in the percentage ever eligible for free school meals improves (although remains significantly different).

Finally, it should be noted that one of the variables that was listed as a proposed control variable is missing for one of the 41 schools in the comparison group. As a result, the effective control group reduces from 41 to 40 schools.

Table 5: Baseline comparison (matched sample)

Variable	Intervention group		Potential comparison group	
School-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Ofsted overall effectiveness outstanding or good	8/14 (0)	57.1	8/41 (0)	56.1
Ofsted quality of teaching outstanding or good	8/14 (0)	57.1	8/41 (0)	57.3
Ofsted leadership and management outstanding or good	9/14 (0)	64.3	9/41 (0)	68.3
School-level (continuous)	**N(missing)**	**Mean**	**N (missing)**	**Mean**
% 5A*-C (including English and maths)	14 (0)	46.1	41 (0)	46.8
Capped points score	14 (0)	221	41 (0)	225
Value added	14 (0)	1,007.4	41 (0)	1,006.8
Teachers with QTS (%)	14 (0)	93.2	40 (1)	94.3
% of pupils eligible for FSM	14 (0)	32	41 (0)	31.3
% of pupils with EAL	14 (0)	28.8	41 (0)	28.4

% of pupils with SEN	14 (0)	10.6	41 (0)	11.6
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Male	1,270/2,441 (0)	52.0	1,189/6,756 (0)	50.9
Ever received FSM	1,151/2,441 (0)	47.2	1,071/6,756 (0)	45.9
Eligible for FSM***	740/2,441 (0)	30.3	633/6,756 (0)	27.1
Pupil-level (continuous)	**N(missing)**	**Mean**	**N (missing)**	**Mean**
IDACI score	2,441 (0)	0.347	6,756 (0)	0.351
Note: Stars for statistical significance: p <= 0.10 (*), p <= 0.05 (**), p <= 0.01 (***).				

Outcomes and analysis

This section presents the analysis of the effect of Achieve Together on pupil-outcomes. All outcomes and analysis are pre-specified and consistent with the Statistical Analysis Plan which is published on the EEF website alongside this report.

To summarise the results: overall, there is no evidence that Achieve Together led to a significant improvement in pupil outcomes. This is because estimates are typically not statistically different from zero as well as being close to zero.

Table 6 presents the estimated impact of Achieve Together on the primary outcome of interest—capped GCSE and equivalents new style point score ('KS4 points') at the end of the two-year intervention. The estimated impact is 0.014 standard deviations, meaning that pupils in intervention schools achieved higher KS4 points scores to pupils in the matched comparison schools by an average of 0.014 standard deviations. This magnitude is classified as a 'low' effect, and in this case is not statistically different from zero. The confidence interval has a lower bound of a negative effect of -0.063 standard deviations and upper bound of 0.091 standard deviations. Importantly, this upper bound is also classified as a 'low' effect on pupil attainment.

Table 6: Primary analysis

	Raw means					Effect size	
	Intervention group		**Control group**				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**n in model (intervention; control)**	**Effect size (95% CI)**	**p-value**
KS4 points score (2016)	2,353 (26)	-0.202 (-0.237;-0.166)	6045 (157)	-0.181 (-0.198;-0.164)	54 (14;40)	0.014 (-0.063;0.091)	0.719

Stars for statistical significance: p <= 0.10 (*), p <= 0.05 (**), p <= 0.01 (***).
Standard errors clustered at school level.
Propensity score weights are applied to the control group.
Outcome is standardised according to the national distribution of scores to have a mean of zero and standard deviation of one.
Pupil-level controls are: season of birth, gender, ever FSM, IDACI score, binary indicator for missing IDACI score, KS2 quartile, and binary indicator for missing KS2 quartile.
School-level controls (2012/2013) are: % achieving five A*–C GCSE and equivalents including English and maths, total average (capped) point score per pupil (GCSEs only), value added measure based on the best eight GCSE and equivalent results, % teachers with Qualified Teacher Status, average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching', latest Ofsted rating for 'leadership and management' at least 'good', % of FSM pupils , % of EAL pupils, and % of pupils with SEN statement or on School Action Plus.

This finding is robust to imposing a common sample for all outcome measures (shown in Appendix F, Table 1) or instead using school-level random effects (equivalent to a multi-level model) to account for the correlation of pupil outcomes within schools (shown in Appendix F, Table 2).

The credibility of these estimates using the set of schools that were interested in Achieve Together to construct the group of well-matched comparison schools is further bolstered by the comparison of this approach with an alternative comparison group for the first cohort of Achieve Together schools. For this first cohort, the second cohort of Achieve Together schools are an ideal comparison group (aside from the small sample size) as they would be hypothesised to have the most similar unobservable (and observable) characteristics to schools in the first cohort of Achieve Together schools (schools that signed up a year earlier). Appendix F, Table 3, compares the estimated impact of Achieve Together after the first year using these two alternative comparison groups of schools. The results give tentative confidence in the suitability of the set of schools that were interested in Achieve Together to construct the group of well-matched comparison schools. The estimated impact on the primary outcome using interested schools to form a well-matched comparison group is not significantly different to the estimate using the second cohort of Achieve Together schools as the comparison group (although the first is slightly negative and the second is slightly positive). This is encouraging, and suggests that unobservable characteristics of schools in the well-matched comparison group are not causing a bias in the estimated impact of Achieve Together, but this is impossible to conclude with the small sample size available.

Table 7 presents the estimated impact of Achieve Together on the secondary outcomes of interest at the end of the two-year intervention. All estimates are small and not significantly different from zero. Note that the negative coefficient for overall absence implies that absence was lower (or better) in intervention schools by 0.024 standard deviations, although, as noted, this difference is not statistically significant. As for the primary outcome, even the upper bound of the confidence interval would be classified as a 'low' impact on attainment and absence.

These findings are robust to imposing a common sample for all outcome measures (shown in Appendix F, Table 1) or instead using school-level random effects to account for the correlation of pupil outcomes within schools (shown in Appendix F, Table 2). It should be noted that the number of control schools in the common sample for all outcomes is lower at 39 schools. This is because pupil absence data is missing for all pupils in one school.

Table 7: Secondary analysis (end of Year 2 of Achieve Together)

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Effect size (95% CI)	p-value
KS4 points score (excluding equivalents) (2016)	2,353 (26)	-0.341 (-0.379; -0.302)	6045 (157)	-0.354 (-0.373; -0.336)	54 (14; 40)	-0.001 (-0.106; 0.104)	0.982
Teach First impact goal measure (2016)	2,353 (26)	-0.326 (-0.364; -0.287)	6045 (157)	-0.332 (-0.351; -0.313)	54 (14; 40)	-0.001 (-0.104; 0.102)	0.982
Overall absence (2016)	2,328 (51)	0.028 (-0.013; 0.069)	5974 (228)	0.019 (-0.001; 0.039)	53 (14; 39)	-0.024 (-0.072; 0.024)	0.315
Stars for statistical significance: p <= 0.10 (*), p <= 0.05 (**), p <= 0.01 (***). Standard errors clustered at school level.							

Propensity score weights are applied to the control group.

Outcomes (aside from absence) are standardised according to the national distribution of scores to have a mean of zero and standard deviation of one.

Teach First impact goal measure represents the capped GCSE point score, excluding equivalents, plus English and maths bonuses.

Overall absence is measured in number of unauthorised absences over the course of the school year and is presented in standardised terms within the estimation sample.

Pupil-level controls are: season of birth, gender, ever FSM, IDACI score, binary indicator for missing IDACI score, KS2 quartile, and binary indicator for missing KS2 quartile.

School-level controls (2012/2013) are: % achieving five A*–C GCSE and equivalents including English and maths, total average (capped) point score per pupil (GCSEs only), value added measure based on the best eight GCSE and equivalent results, % teachers with Qualified Teacher Status, average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching', latest Ofsted rating for 'leadership and management' at least 'good', % of FSM pupils, percentage of EAL pupils, and % of pupils with SEN statement or on School Action Plus.

Table 9 presents the estimated impact of Achieve Together on the secondary outcomes of interest at the end of the first year of the intervention. The well-matched comparison group of schools is identical here to that used in the primary analysis. The outcomes do, however, relate to the cohort of pupils at these schools that are one year older. The scale of missing pupil attainment data is very similar to the main analysis, but the exact numbers are slightly different.

In contrast to the estimated impact on attainment at the end of intervention, there is some suggestion that Achieve Together had a negative and significant impact on attainment after the first year of the programme, though this is still a small effect and is only statistically significant at the 10% level. Estimates suggest that pupil attainment was lower in intervention schools by around 0.07 standard deviations relative to well-matched comparison schools. Although these estimates are small, they may suggest that Achieve Together had a negative impact on pupil attainment while programmes and collective impact projects were being established. The improved trajectory of attainment in intervention schools may suggest that the positive impact of Achieve Together will become evident in the longer-term, but a subsequent analysis of future cohorts of pupils is needed to determine this. These results are robust to using a prior measure of pupil attainment at KS2 based on teacher assessment rather than test scores, which is explored as over 40% of pupils in our sample were affected by the boycott of KS2 tests in 2010 (see Appendix F, Table 4).

Table 8: Secondary analysis (end of Year 1 of Achieve Together)

| | Raw means | | | | Effect size | | |
| | Intervention group | | Control group | | | | |
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Effect size (95% CI)	p-value
KS4 points score (2015)	2,418 (33)	-0.222 (-0.256; -0.189)	6,372 (144)	-0.158 (-0.174; -0.141)	54 (14; 40)	-0.067 (-0.147; 0.013)	0.092*
KS4 points score (excluding equivalents) (2015)	2,418 (33)	-0.349 (-0.386; -0.312)	6,372 (144)	-0.338 (-0.356; -0.320)	54 (14; 40)	-0.068 (-0.157; 0.020)	0.12
Teach First impact goal measure (2015)	2,418 (33)	-0.326 (-0.363; -0.290)	6,372 (144)	-0.313 (-0.332; -0.295)	54 (14; 40)	-0.067 (-0.145; 0.012)	0.089*

| Overall absence (2015) | 2,394 (57) | -0.047 (-0.081; -0.013) | 6,281 (235) | -0.046 (-0.064; -0.029) | 54 (14; 40) | 0.007 (-0.037; 0.050) | 0.763 |

Stars for statistical significance: p <= 0.10 (*), p <= 0.05 (**), p <= 0.01 (***).

Standard errors clustered at school level.

Propensity score weights are applied to the control group.

Outcomes (aside from absence) are standardised according to the national distribution of scores to have a mean of zero and standard deviation of one.

Pupil-level controls are: season of birth, gender, ever FSM, IDACI score, binary indicator for missing IDACI score, KS2 quartile, and binary indicator for missing KS2 quartile. KS2 quartile is based on teacher assessment where test scores are missing due to KS2 boycott in 2010.

School-level controls (2012/2013) are: % achieving five A*–C GCSE and equivalents including English and maths, total average (capped) point score per pupil (GCSEs only), value added measure based on the best eight GCSE and equivalent results, % teachers with Qualified Teacher Status, average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching', latest Ofsted rating for 'leadership and management' at least 'good', % of FSM pupils, % of EAL pupils, and % of pupils with SEN statement or on School Action Plus.

Table 10 shows that the impact of Achieve Together on the primary outcome is not stronger for the group of schools classified as complying with the original design of Achieve Together. 'On-treatment' is defined as having at least one participant from the school on the senior leadership training programme (NPQSL) over the two-year programme. This applied to 10 of the 14 intervention schools.

Table 9: Primary analysis (on-treatment)

| | Raw means | | | | Effect size | | |
| Outcome | Intervention group | | Control group | | | | |
	n (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)	n in model (intervention; control)	Effect size (95% CI)	p-value
KS4 points score (2016)	1,684 (18)	-0.217 (-0.258; -0.176)	6045 (157)	-0.181 (-0.198; -0.164)	50 (10; 40)	-0.000 (-0.095; 0.095)	0.999

Stars for statistical significance: p <= 0.10 (*), p <= 0.05 (**), p <= 0.01 (***).
Standard errors clustered at school level.
Propensity score weights are applied to the control group.
Outcome is standardised according to the national distribution of scores to have a mean of zero and standard deviation of one.
'On-treatment' is defined as having at least one participant from the school on the senior leadership training programme (NPQSL) over the two-year programme.
Pupil-level controls are: season of birth, gender, ever FSM, IDACI score, binary indicator for missing IDACI score, KS2 quartile, and binary indicator for missing KS2 quartile.
School-level controls (2012/2013) are: % achieving five A*–C GCSEs and equivalents including English and maths, total average (capped) point score per pupil (GCSEs only), value added measure based on the best eight GCSE and equivalent results, % teachers with Qualified Teacher Status, average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching', latest Ofsted rating for 'leadership and management' at least 'good', % of FSM pupils, % of EAL pupils, and % of pupils with SEN statement or on School Action Plus.

Table 11 shows the impact of Achieve Together on the primary outcome for sub-groups of pupils as pre-specified. The classification of ever-eligible for free school meals ('ever-FSM') is taken from the NPD, and 'low prior attainment' is defined as not achieving at least level 4 at Key Stage 2 in each of maths, English, and (teacher assessed) science.

There are no significant differences between sub-groups of pupils, suggesting that the programme was not differentially effective for different groups of pupils.

Table 10: Primary outcome (sub-group analysis)

| Outcome | Raw means | | | | Effect size | | |
| | Intervention group | | Control group | | | | |
	n (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)	n in model (intervention; control)	Effect size (95% CI)	p-value
Ever-FSM	1,100 (11)	-0.368 (-0.421; -0.316)	2,835 (56)	-0.360 (-0.386; -0.335)	54 (14; 40)	0.003 (-0.077; 0.083)	0.933
Never-FSM	1,253 (15)	-0.056 (-0.103; -0.009)	3,210 (101)	-0.025 (-0.048; -0.003)	54 (14; 40)	0.006 (-0.074; 0.087)	0.873
Low prior attainment	722 (10)	-0.754 (-0.809; -0.699)	1,885 (35)	-0.714 (-0.740; -0.687)	54 (14; 40)	-0.030 (-0.121; 0.060)	0.504
High prior attainment	1,403 (5)	0.149 (0.112; 0.185)	3,577 (121)	0.166 (0.148; 0.183)	53 (14; 39)	0.006 (-0.070; 0.082)	0.871

Stars for statistical significance: p <= 0.10 (*), p <= 0.05 (**), p <= 0.01 (***).

Standard errors clustered at school level.

Propensity score weights are applied to the control group.

Outcome is standardised according to the national distribution of scores to have a mean of zero and standard deviation of one.

'Low prior attainment' is defined as not achieving at least level 4 at Key Stage 2 in each of maths, English and (teacher assessed) science. KS2 prior attainment is missing for 902 pupils, which includes some pupils for whom GCSE attainment was also missing. As a result, the number of missing pupil observations is different compared with previous analysis.

Pupil-level controls are: season of birth, gender, ever FSM, IDACI score, binary indicator for missing IDACI score, KS2 quartile, and binary indicator for missing KS2 quartile.

School-level controls (2012/2013) are: % achieving five A*–C GCSEs and equivalents including English and maths, total average (capped) point score per pupil (GCSEs only), value added measure based on the best eight GCSE and equivalent results, % teachers with Qualified Teacher Status, average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching', latest Ofsted rating for 'leadership and management' at least 'good', % of FSM pupils, % of EAL pupils, and % of pupils with SEN statement or on School Action Plus.

Cost

The total cost of Achieve Together from a schools' perspective depends on the cost of each programme and the total number of participants per programme. Table 12a provides an overall summary of our cost analysis by showing the total cost to the school per participant on each programme and the total staff days required per participant (including how much of this was within the school). The final column then shows average number of teachers per school on each programme, which is a crucial ingredient for the rest of the cost analysis. Note that these averages are slightly different to those shown in Table 3 as the cost analysis includes two extra schools who took part in the Achieve Together, but were not included in the impact evaluation as they did not meet the pre-specified criteria of signing a memorandum of understanding.

Table 112a: Summary of costs and staff time

Item	Total cost to the school per participant	Total staff days per participant (of which in school)	Average number of teachers per school who did this course.
Teach First	£8,100	48.8 (17.3)	4.3
TL NPQML	£1,000	3.6 (0)	4.6

TL Fellows	£4,000	22.9 (4.3)	1.1
NPQSL	£1,500	7.5 (0)	1.6
Future leaders	£0	51.1 (28.2)	0.2

The rest of the analysis in this section looks at the costs and staff time in more detail. We provide the average total cost per pupil (and average total cost per pupil per year assuming a three-year intervention) based on the number of participants in intervention schools (as shown in Table 12a above). For the benefit of schools thinking of adapting such a programme to their own needs, we also report the cost per participant per year for each programme.

Table 12 shows that the average total cost of Achieve Together for intervention schools was around £50,000 over three years (assuming no new participants in the third year). The variation in total cost over three years was between £17,000 and £75,000. Divided by the average number of pupils in intervention schools (931), the average total cost over three years is £49.40 per pupil, and £16.47 per pupil per year.

Table 12: Cost per pupil per year from the schools' perspective

Item	Cost per participant per year	Average total cost over 3 years	Average total cost per pupil over 3 years	Average total cost per pupil per year
Teach First	£4,050.00	£34,425.00	£36.98	
TL NPQML	£1,000.00	£4,625.00	£4.97	
TL Fellows	£2,000.00	£4,500.00	£4.83	
NPQSL	£1,500.00	£2,437.50	£2.62	
Future Leaders	£0.00	£0.00	£0.00	
Total		£45,987.50	£49.40	£16.47
Note: Costs per year are those provided by Teach First, Teaching Leaders, and Future Leaders per participant per year, from the school's perspective. Teach First, TL Fellows, and Future Leaders are two-year programmes and costs are therefore applicable for two years. The average total cost per programme over three years multiplies the number of participants per school over three years (assuming no new participants in the third year) by the cost per year, and takes the average across all intervention schools. The average total cost per pupil over three years divides this cost by the average number of pupils in intervention schools (931). The average total cost per pupil per year then divides this total by three for the average cost per year.				

Teach First, TL Fellows, and Future Leaders receive government funding per participant (even in the absence of being part of Achieve Together). Table 13, therefore, replicates Table 12 using the total cost to schools and government per participant per year. The average total cost of Achieve Together for government and intervention schools was around £215,000 over three years (assuming no new participants in the third year). The variation in total cost over three years was between £49,000 and £367,700. Divided by the average number of pupils in intervention schools (931), the average total cost over three years is £230.52 per pupil, and £76.84 per pupil per year.

Table 13: Cost per pupil per year from the schools' and government's perspective

Item	Cost per participant per year	Average total cost over 3 years	Average total cost per pupil over 3 years	Average total cost per pupil per year
Teach First	£18,550.00	£157,675.00	£169.36	
TL NPQML	£1,000.00	£4,625.00	£4.97	
TL Fellows	£18,000.00	£40,500.00	£43.50	
NPQSL	£1,500.00	£2,437.50	£2.62	

Future Leaders	£25,000.00	£9,375.00	£10.07	
Total		£214,612.50	£230.52	£76.84

Note: Costs per year are those provided by Teach First, Teaching Leaders, and Future Leaders, per participant per year, from the schools' perspective. Teach First, TL Fellows, and Future Leaders are two-year programmes and costs are therefore applicable for two years. The average total cost per programme over three years multiplies the number of participants per school over three years (assuming no new participants in the third year) by the cost per year, and takes the average across all intervention schools. The average total cost per pupil over three years divides this cost by the average number of pupils in intervention schools (931). The average total cost per pupil per year then divides this total by three for the average cost per year.

In line with EEF guidance, staff time costs are not included in the above cost estimates but are presented separately in Table (for staff time in school hours) and Table 6 (staff time outside school hours). These totals only account for involvement in the individual programmes. They exclude staff time for involvement in the collaborative impact project as there was a low response rate to the survey (four schools) which aimed to give an indication of this. The average amount of staff time spent on Achieve Together programmes across each school (including both in-school and out-of-school hours) was around 270 days over three years (again assuming no new participants in the third year). The variation in total staff time over three years was between 74 days and 395 days. Divided by the average number of pupils in intervention schools (931), the average total staff time taken over three years is 0.3 days per pupil, and 0.1 days per pupil per year.

Table 15: Staff time (in-school hours)

Item	Staff time (days) per participant per year	Average total staff time (days) over 3 years
Teach First	8.7	73.7
TL NPQML	0.0	0.0
TL Fellows	2.1	4.8
NPQSL	0.0	0.0
Future Leaders	14.1	5.3
Total		83.7

Note: The staff time required per programme is provided by Teach First, Teaching Leaders, and Future Leaders, per participant. Teach First, TL Fellows, and Future Leaders are two-year programmes and staff time is therefore applicable for two years. One day means 7.5 hours of staff time. For both the Teach First and TL Fellows programmes, quoted time spent on in-school hours is likely to be an upper bound and time spent on out-of-school hours a lower bound, but the total is approximately correct. The average total staff time per programme over three years multiplies the number of participants per school each year (assuming no new participants in the third year) by the total staff time, takes the total across all three years, and then calculates the average of this three-year total across all intervention schools.

Table 16: Staff time (out-of-school hours)

Item	Staff time (days) per participant per year	Average total staff time (days) over 3 years
Teach First	15.7	133.7
TL NPQML	3.6	16.7
TL Fellows	9.3	21.0
NPQSL	7.5	12.1
Future leaders	11.5	4.3
Total		187.8

Note: The staff time required per programme is provided by Teach First, Teaching Leaders, and Future Leaders, per participant. Teach First, TL Fellows, and Future Leaders are two year programmes and staff time is therefore applicable for two years. One day means 7.5 hours of staff time. For both the Teach First and TL Fellows programmes quoted time spent on in school hours is likely to be an upper bound and time spent on out of school hours a lower bound, but the total is approximately correct. The average total staff time per programme over 3 years multiplies the number of participants per school each year (assuming no new participants in the third year) by the total staff time, takes the total across all three years, and then calculates the average of this three-year total across all intervention schools

Process evaluation

This section reports on findings from the process evaluation. It begins by examining implementation of Achieve Together, including the key barriers and facilitators to successful implementation. It then goes on to assess the fidelity of the intervention and perceived impacts. The section ends by summarising formative findings, about how Achieve Together could be developed in the future.

Implementation

This section explores the implementation of key components of the Achieve Together model, examining the barriers and facilitators to successful implementation and how the initiative was received by participating schools. It is structured around the key components of the initiative, drawing on the theory of change developed by the Achieve Together project team (Appendix D). We begin by considering:

1. recruitment to the pilot and clarity of aims and purpose.

We then consider Achieve Together activities:

2. needs analysis process;
3. collaborative planning meetings, looking at
   a. facilitators of effective implementation, and
   b. barriers to effective implementation;
4. collective impact initiatives, looking at
   a. facilitators of effective implementation, and
   b. barriers to effective implementation; and
5. support activities.

1. Recruitment to the pilot and clarity of aims and purpose

Achieve Together was devised to support leadership development and collaboration within schools to foster whole-school improvement and ultimately improve outcomes for children. The initiative aimed to do this by subsidising leadership development training for teachers in middle and senior leadership roles, and by placing graduates into schools via the Teach First initiative. Alongside this investment in CPD, Achieve Together offered a range of support to facilitate collaboration and alignment across these programmes, with a specific focus on participants working collaboratively on a school improvement impact initiative:

> *'The kind of key difference in terms of Achieve Together would be that rather than just looking at their individual impact initiatives, they work together with a collective in the school to work on a targeted intervention'* (AT project manager, Year 1).

The evaluation found that the collaboration and alignment of leadership programmes (and in particular the expectation that participants would work together on a school impact initiative) was not well understood by schools when they signed up to the programme. Schools largely conceptualised the initiative as a way of accessing a range of CPD, and the Achieve Together project manager acknowledged that this was the focus of initial recruitment to the pilot. This lack of clarity regarding aims of Achieve Together and how it worked alongside the individual leadership programmes, was apparent across the full range of staff participating in the leadership programmes:

> *'I've never truly understood what the aim is … I can remember even when we went to those meetings during our training, a lot of people came out of it going, 'I still don't understand what it is', because it's not really clear what they're trying to do. It's the bringing together of staff from different levels, you can see the benefits in that, but why, to what end? And I think it's—I realise*

now it's a case of your school decides to what end, your school decides what it is you want to do, what it is you need to do. I don't think that's made that clear from the outset that your school designs what you need and that kind of thing' (Teach First participant, Year 1, School 5).

Ultimately, this lack of clarity in terms of the initiative's aims and key components affected implementation in so far as staff in case-study schools were unclear about how the programme was intended to work. In hindsight, the project manager reflected, much more emphasis should have been placed on the collaborative elements of the pilot during this initial recruitment stage:

'If we were to do the recruitment now, so much of it would be about identifying collaborative ways of working towards school improvement and continuous CPD, rather than talking about just alignment of CPD, vertical alignment of CPD, because I think that's what the initial offer was very much talking about vertical alignment of CPD across the school' (project manager, Year 2).

This view was supported by staff in the case-study schools participating in the initiative who felt that detail on how Achieve Together was intended to work needed to be made much more explicit at the outset to ensure its success:

'I think basically if they'd come in and said, you know, "Whoever takes on these courses, you will be part of the bigger picture within your school … you will need to co-ordinate with other people to make sure that your initiatives work together to ..."—yeah, just that bit of information was missing and by the time we got the information, people had already started down their project road ... [That] kind of killed a lot of kind of enthusiasm for the bigger project, I think, because we thought, "Well, why didn't we know that? What do you expect us to do now?"' (NPQSL participant, Year 1, School 5).

2. Needs analysis process

To support the alignment of the leadership programmes with school improvement goals, the Achieve Together project team carried out a needs analysis process in each school at the start of the first year of the pilot. This typically involved meeting with the headteacher and Achieve Together in-school lead; and reviewing key documents including school improvement plans (SIPs) and self-evaluation forms (SEFs) alongside school data. The aim of this process was to identify the three areas that would become the focus of future collaboration across the leadership programmes, and to get the buy-in and commitment from the headteacher.

This process received mixed feedback from participating schools. In some instances, schools reported finding it useful, valuing it as a form of external 'QA' of their school development plans and priorities. In other cases, the process was viewed as time consuming and was felt to be more about 'information giving' rather than as a process that provided useful insights or added value to their existing assessments:

'The people that were coming to us to do the needs analysis took an approach that felt very much like they believed—this was our perception—that we wouldn't necessarily be able to identify our own needs and be able to identify our own priorities … it was a little bit difficult because they didn't really say anything we didn't already know. But it took a considerable amount of time to kind of go round the houses and get back to the same point. And it was quite a frustrating experience so it didn't set things up particularly well' (in-school lead, Year 2, School 6).

3. Collaborative planning meetings

To support alignment and collaboration across the leadership programmes with the aim of working towards a whole-school improvement project, Achieve Together held collaborative planning meetings in each of the participating schools. These meetings were typically facilitated by the Achieve Together project manager, attended by the headteacher, the Achieve Together in-school lead and all staff participating in the leadership programmes and lasted between 2.5 and 3 hours. Schools identified the following strengths and limitations to this approach:

3.a Facilitators of implementation

Where the collaborative planning worked well, participants reported the following elements being useful:

- Collaboration across staff
  Staff valued the opportunity the meetings gave for staff at different stages of their careers to collaborate in a process that was looking at school improvement strategies:

  'Well it was particularly powerful for the younger staff to be able to sit down and listen to more experienced members of staff thinking about school … It was a real eye-opener to think about the strategic planning that goes on' (in-school lead, Year 1, School 5).

- Effective planning tools
  Staff also praised the use of the 'Business Canvas Model', a tool adapted by AT to plan school improvement initiatives. The tool was felt to aid good quality discussions, help generate ideas, and focus on practical implementation:

  'I really liked the business model and I think that's something we'll use again at SLT when we're talking through issues and planning. I think it was a really effective way of putting together your improvement plan ... it led to a real quality discussion. There were really tangible outcomes at the end' (in-school lead, Year 1, School 2).

3.b Barriers to effective implementation

This section summarises the main barriers to effective implementation. These related to timing, resource requirements, and delays in receiving feedback.

- Timing
  The timing of the meetings (which did not take place until the spring term of the first year of the pilot) was felt to undermine their value because participating staff had made significant progress on identifying their individual impact initiatives before this point and were not in a position to change these projects to align more strategically with the plans generated by these meetings:

  'I think it was April, though, and that's been the drawback really because all these good things have happened so, so very late … I mean, if we'd had this in the autumn term it might have made an impact. I can't hand on heart say there's been no impact whatsoever because it's all happened so late in the day' (in-school lead, Year 1, School 5).

  The Achieve Together project manager reflected that it was logistically difficult to find dates where all staff could be brought together, and the Achieve Together team also acknowledged that it struggled to resource the process with limited staff. Delays to the recruitment of schools because of the change in evaluation methodology from RCT to matched comparison group design may also have contributed to this issue:

> *'Even the first schools that we went into, it was still after most of the participants on programmes had set their individual impact initiatives, and some schools felt quite frustrated about that because they weren't able to tally the two together'* (project manager, Year 1).

This issue fundamentally undermined the effectiveness of this planning process as it came too late to be of use for a number of participating schools.

- Resource intensive
  Schools reflected on the expense of absenting so many staff from lessons to participate in the meeting. Consequently, some schools held the meetings after school to minimise the costs of providing supply cover:

> *'Yeah, we were determined we couldn't spend any time during the school day. That wasn't going to be on … there would be six Teach First, two middle leaders … plus some senior team there that want to get involved—12, 13, 15 people at the school … that's 15 afternoons we couldn't cover. That's chaos around the school potentially because that's so many people and incredibly expensive'* (in-school lead, Year 1, School 5).

In other cases, schools paid for cover to allow staff time within the school day to attend this meeting, but acknowledged that this was a costly approach.

- Delays in receiving feedback
  Schools reported that Achieve Together were slow to provide feedback from the meetings, with some not receiving copies of outputs from the process until the end of the first year:

> *'I was under the impression that we would have a little bit more input and support in this … we did have a [planning meeting] in February … What they did in the meeting was quite useful, in terms of ways of planning and strategy, but I still haven't had any feedback from it [in April]'* (in-school lead, Year 1, School 4).

These delays were felt to limit the momentum built during the planning sessions and undermine their value.

4. Collaborative impact initiatives

At the heart of the Achieve Together model is the concept of increased collaboration between staff to affect whole-school improvement. Using the leadership programmes as a vehicle to promote this collaborative model, Achieve Together aimed to encourage collaboration by aligning the impact initiatives (that form part of the leadership programmes) to feed into a whole-school project. 'Impact initiatives' were projects carried out by participating staff in areas of school improvement that related to their areas of responsibility. They formed part of the assessment requirements for the Flagship, TL Fellows, NPQML, and NPQSL programmes (with some variation in focus depending on the programme)[12] and typically involved analysis of school data to identify an area for improvement, delivery of an intervention, and an assessment of impact.

The process evaluation found limited evidence of alignment and collaboration across the leadership programmes and this section explores views on the feasibility of the model and the barriers and facilitators to implementation reported by participating schools.

<u>Feasibility of collaboration across the leadership programmes</u>

---

[12] For example, on the Future Leaders flagship programme this involves a 'whole-school impact initiative', while on the Teaching Fellows programme this is a two-year impact initiative focused on its area of leadership responsibility.

Schools reflected that the concept of collaboration across the leadership programmes was a positive one. In particular, staff highlighted the benefits of learning from each other, sharing ideas, and the potential for greater collaboration to reduce workload:

> *'In terms of aligning projects it's a really good idea and it makes absolute sense, because then you're working with lots of people, people from different departments, not just your own to make a bigger impact in school … I do think it would be really beneficial, especially if you could see the outcomes by the end of it all'* (TL Fellows, Year 1, School 5).

However, while the general concept was received positively, three caveats were raised:

- The role of Teach First participants
  Questions were raised over the extent to which Teach First participants could effectively engage in Achieve Together because of the demands of learning to teach and achieve qualified teacher status:

  > *'I think at the time I was so overwhelmed with everything else that was going on, it almost felt like "Oh my gosh, we've got something extra to do now". Like not only have I actually got to learn how to teach, I've also got to do this kind of extra thing'* (Teach First participant, Year 1, School 5).

  In recognition of this concern, the Achieve Together project manager reflected that a 'light touch' approach should be taken to the participation of Teach First staff, potentially including them in planning meetings but limiting expectations that they would lead the delivery of impact initiatives. Schools also suggested that Teach First candidates in their second year of teaching, or potentially those who had qualified and were in their third year, would be better placed to contribute to Achieve Together.

- Type of impact initiative
  To foster collaboration, schools reflected on the importance of selecting a project that had a 'whole-school' focus and was relevant to staff across faculties and roles. While this was feasible when staff had sufficient time to plan the project, some held the view that it was not always desirable to select an initiative of this kind and highlighted the value of individual impact initiatives focusing on subject-specific areas of practice with perhaps less scope for collaboration.

- Compatibility with staff roles
  A key consideration for the feasibility of the collaborative model was the extent to which the impact initiatives were compatible with participating staff interests and roles. This was felt to be a key challenge for the model because the relevance of the project was thought to be essential for staff engagement. In cases where the project was viewed as an 'add-on' that did not speak to their existing roles and responsibilities, staff struggled to find time to contribute because of their existing commitments and priorities:

  > *'The potential is phenomenal but … Teach First are going through a really intensive programme … The Fellows are doing a very intensive programme with a lot of expectations in social time … And then you've got people like myself—the NPQSL—where it was almost an add-on to my role, but I've turned it into something so that it linked into my role, as did my colleague because otherwise there's no way it would work. And then you're then trying to get those three groups together to then be involved in running something whole-school. The potential is really high but there's a lot of things there which are going to make it difficult to work'* (in-school lead, Year 1, School 4).

4.a Facilitators to effective implementation
In implementing this collaborative model of school improvement, the following facilitators were identified:

- Timing of Achieve Together activities
  In schools where the Achieve Together planning meetings preceded the selection of impact initiatives by staff participating in the leadership programmes, this provided greater scope to align projects and work more collaboratively. In one school, for example, the NPQML programme did not start until January of the first year of the pilot which meant the collaborative planning session could inform project selection and also gave staff the opportunity to explore how the projects could tie in together.

- Leadership programme structure
  Schools that had some success in aligning projects across the leadership programmes valued the structure and impetus this gave to project implementation. In particular they valued the increased involvement of staff at different stages of their careers that was fostered by the model:

  'I think it's changed the dynamic in the sense that its given them that extra drive with a named person responsible for them … So I suppose in a way, instead of it being so top driven, maybe, which potentially could have happened if it was senior team led and in the SIP, I think it's had a much more democratic drive to it. And I think probably it's brought more staff on board because the whole cohort of people were involved in the joint planning session' (in-school lead, Year 1, School 2).

- Clear focus and leadership from SLT
  To foster and encourage collaboration, schools reflected that senior leaders played a key role in identifying the areas for school improvement and providing a clear direction for staff. The Achieve Together project manager viewed the engagement of senior leaders as critical to the success of the programme:

  'The in-school lead engagement is really important, but equally important is how well the in-school lead is supported by the head … Where the in-school lead has been engaged but lacked not necessarily support, but buy-in from the head teacher, it's been much more difficult for them to actually create the space within the wider teaching and learning strategy to implement something like this' (Achieve Together project manager, Year 2).

- Staff within faculty
  When staff collaborating on initiatives were in the same faculty this was felt to facilitate joint working because there was greater scope for informal catch-ups and closer working ties. Staff reflected that projects that were running across faculties were feasible but required more time and effort to co-ordinate.

Case example 1: collaboration across leadership programmes (School 2)

In this school, a senior member of staff participating in the NPQSL with Future Leaders undertook an impact initiative that focused on improving the school's approach to fostering students' attitudes to learning. Her aim was to evaluate what they already had in place, identify ways in which to improve it, and then roll out these improvements across the school. Alongside her, four members of staff participating in the NPQML with Teaching Leaders also undertook impact initiatives that looked at various elements of Attitudes to Learning within their own faculties which contributed and fed into her project:

'I've got some members of staff who are not on the NPQML who are trialling different bits … who are dedicated to looking at baseline data, trialling strategies, evaluating the data afterwards, and kind of essentially it just it helps me just to pull in and look at actually what does and doesn't work. And it's been really useful as well for them bouncing ideas off each other so they've tried different ideas in different year groups or the same idea in different year groups … And that's been really good.'

As a result of the project, changes were made to the frequency that students received feedback on their attitudes to learning, from termly to half-termly. A simplified measure of attitudes that students found easier to understand was also rolled out and the approach was embedded across the school with custom made exercise books rolled out in the second year of the pilot to encourage staff and students to reflect on Attitudes to Learning as an integral part of assessment and marking.

Case example 2: collaboration across leadership programmes (School 6)

This school had a whole-school focus on improving the attainment of students eligible for the pupil premium. A member of staff participating in the NPQSL with Future Leaders undertook an impact initiative to improve the marking and feedback strategies used across the school, with a particular focus on pupil premium students.

Another project was being carried out by a staff member participating in the NPQML with Teaching Leaders. This project also focused on improving attainment, focusing specifically on improving the maths attainment of a group of White British boys who had been identified as underachieving. While this project was testing the value of a peer mentoring approach, it was also possible for the staff member to trial and test some of the feedback and marking strategies being developed by the staff member on the NPQSL and contribute to the wider whole-school initiative of improving marking and feedback strategies.

4.b Barriers to effective implementation
Where implementation was less successful, the following barriers were identified:

- Timing of Achieve Together planning meetings
  As previously discussed, the delays in holding the collaborative planning meetings until the spring term of the first year limited the extent to which some schools were able to collaborate across initiatives. This was further exacerbated by delays in receiving feedback from these meetings. By this point individual impact initiatives had been agreed and it was too late, in a lot of cases, to find ways in which to align activities.

- Communication and clarity of concept
  The lack of clarity around the purpose of Achieve Together and how it might work in practice (discussed in relation to recruitment to the pilot earlier) also impacted on the extent to which schools were able to collaborate across initiatives. In particular, the communication of the aims and purpose of the model beyond the in-school lead to the staff participating in the individual programmes was

not effective. Staff reported that they were unclear about how this would work in practice, and by the time the collaborative element was made clear, it was too late:

'That has always been the case that we've never fully understood how that's worked. So, when I came back and spoke to people here [at school], they were not aware of that either … everybody was working on different projects, we knew that bit, but we didn't know about how it was all meant to feed in. So that was a bit of news, but people were already down the project road, had already gone and started projects … so we couldn't change anything, so we've just carried on as best we can' (NPQSL participant, Year 1, School 5).

- Leadership programme lengths
Although the pilot was for two years, the various leadership programmes varied in length. Some staff reflected that the level of collaboration weakened after some staff completed their individual impact initiatives at an earlier stage than others. This feedback suggests that more consideration should be given to the compatibility of the leadership programmes (in terms of length and structure) to maximise opportunities for collaboration.

- School capacity
Achieve Together was designed as a 'self-led' model of school improvement with participating schools expected to drive the project forward with support from the leadership programmes and the Achieve Together project team. In practice, the project manager reflected that schools had varying capacity to take on this role and this inevitably impacted on the extent to which they were successful in working collaboratively on an impact initiative. Examples of factors that undermined the capacity of schools to engage included instability in senior leadership teams, financial deficits, and Ofsted judgments that focused staff attention elsewhere and limited their capacity to engage with Achieve Together:

'I think our school this year has been about Ofsted and we had two "required improvements", so this was going to be our make or break Ofsted, and so all of the school's energy was going into making sure that was right and I think Achieve Together just didn't have time … and I think as well because it was SLT who were watching Achieve Together, they had so many things to do in addition to that. So, I think it just happened very, very late in the year … I think that's basically what happened' (Teach First participant, Year 1, School 5).

- Staff changes
While the in-school leads remained fairly consistent across the two-year pilot period, there were examples of sick leave and maternity leave that impacted on communication between the Achieve Together team that led to delays in implementation.

5. Support activities

As the pilot progressed, the Achieve Together project team developed some additional materials to support participating schools further. The development of these resources was in response to the 'self-led' nature of the Achieve Together model and a desire by the Achieve Together project manager to explore ways to build capacity for school improvement in a format that was scalable:

'We've concentrated on activities and products where we felt we could play that influencing role without actually having to go in and visit the school' (Achieve Together project manager, Year 2).

Two key developments in the second year of the pilot were a package of online CPD resources and a series of webinars on aspects of school improvement:

- CPD resources
Toward the end of the first year of implementation, the Achieve Together project team worked with a school improvement charity that had developed a set of online tools and resources known as a 'route map'. The resources, which included research evidence, CPD materials, and elements of action research, were tailored to focus on three areas of school improvement that the Achieve Together needs analysis process had helped identify – formative assessment, coaching and learning behaviours.

From a school perspective, there was mixed feedback on the value of the route map. In one school, for example, they embedded it into their regular staff CPD programme and it was being used by staff across the school at all stages of their careers. Staff who used the route map spoke positively about the breadth of the resources, its accessibility, and the way in which it encouraged discussion and reflection:

> *'So we're encouraging them to go and look at the research, pull things out … to talk about it. So it's a really quick and easy way of getting research out there amongst the staff, so that's been really positive, and we'll continue to work on that next year, I think'* (in-school lead, Year 2, School 2).

Where the route map had less impact, this was for a number of reasons. In some instances staff did not have the capacity to engage with it effectively:

> *'It's another thing on top of like the NPQML, on top of all your marking, on top of overseeing learning and whatever. It's something we're aware that we should be doing but in a realistic world we haven't'* (NPQML participant, Year 2, School 2).

In other cases staff were unaware of it, suggesting that lines of communication between Achieve Together in-school leads and the wider cohort of Achieve Together participants were not always robust.

- Webinars
A series of five webinars were held throughout the second year of the pilot that were open to all Achieve Together participants. Typically 45 minutes in length, the webinars were led by a range of experts and covered:

  o leadership and coaching;
  o evidence-based approaches to self-improving schools;
  o evidence-based practice and school improvement;
  o formative assessment; and
  o the Mental Toughness Questionnaire.

Held at 4pm, the Achieve Together project team reflected that 'live' engagement with the webinar series was relatively low, and feedback from schools confirmed that it was difficult for staff to find time to attend the online events because of a large number of other school commitments. However, the webinars were archived so that teachers could access them on demand, but not all staff interviewed were aware of this. In general, staff were positive about using webinars to support CPD, reducing the time required to travel to face-to-face meetings, although some expressed a preference for face-to-face training.

Fidelity

Achieve Together was not a manualised intervention with prescribed programme content. Indeed, the scope and format of Achieve Together changed and developed over the course of the two-year pilot

with additional features introduced in the second year to explore their efficacy. Given this flexibility it is difficult to make a judgement on fidelity to the programme. However, across the participating schools there was variation both in terms of the number of staff participating in the leadership programmes (from 5 to 23) and in terms of which leadership programmes were taken-up. This variation may impact on the fidelity of the programme by limiting the scope for collaboration in schools with fewer staff participating.

The process evaluation also found limited evidence of collaboration across impact initiatives, indicating that fidelity to this key component of the model was not strong.

Outcomes

Achieve Together aimed to foster collaboration and strategic alignment between leadership development training and whole-school improvement priorities, with the ultimate goal of improving pupil attainment. This section reports on the perceived outcomes from the perspective of participating schools.

Staff outcomes

A range of positive impacts were identified from participation in the individual leadership programmes that were the foundation of the Achieve Together programme. Staff described thinking more strategically, developing their confidence, improving their management skills, and using these skills to support other staff and contribute to school improvement:

> *'It's one of the best programmes I could've ever been on, to be honest. Having a coach talking to me, getting me to think in a different kind of mind-set has been really, really valuable ... And I've been using a lot of those coaching and mentoring techniques that I've gained on the course with the staff that I work with. So they have been really fantastic to get people to really think about how they're developing their own CPD, how they are leading in their classroom in that context'* (TL Fellows participant, Year 2, School 5).

Schools welcomed the opportunity Achieve Together provided to access these programmes and invest substantially in staff development at all levels. In some instances, this was the first time they had invested in leadership development on a large scale, and the collective impact of investing in staff across the school was felt to be positive:

> *'It adds a sense of pace, urgency, sense of a drive to the participants, and the more people who have got that feeling the better'* (in-school lead, Year 2, School 5).

Some internal promotions were also attributed to the skills developed on the leadership programmes, and this was viewed positively, although some concerns were raised over the potential for staff retention to be negatively affected because more skilled staff became more employable elsewhere:

> *'I think one of the pieces of data that we may or may not be able to get hold of is also around the effect on retention of individual leadership development. So we have some anecdotal examples of where teachers that are working in more challenging schools are kind of moving on because they've achieved a kind of greater level of autonomy as a result of the work that they've been doing in the schools. So for schools where this was … a heavy investment in CPD for the first time in some years, we're just keen to see what the net effect has been … on retention'* (Achieve Together project manager, Year 2).

School improvement

Schools were divided on the extent to which they felt the Achieve Together programme had contributed to school improvement (beyond the impact of the individual leadership programmes). For schools that

succeeded, in some measure, in aligning impact initiatives with school improvement goals and capitalising on the resources offered by Achieve Together, positive outcomes included:

- An increase in school improvement impetus
  By embedding school improvement goals within the impact initiatives, the programme was felt to give added impetus to school improvement by focusing staff on the improvement priorities:

  'I think the fact that Achieve Together allows you to have NPQSL, NPQML, Teaching Fellows, and they can all pick up on those strands, that's really, really helpful' (in-school lead, Year 2, School 2).

- An increase in reflective practice
  Involving a wider group of staff in planning school improvement, and capitalising on resources provided by Achieve Together, meant that staff developed more reflective practice and this was viewed positively by senior leaders:

  'It has made us think about what we do, and who we include within that … And I suppose if there's anything that Achieve Together, probably, wanted schools to do, it was that, to evaluate, to look at what you're doing and why, and is it the best way for you to do it, and what's helping you inform that' (in-school lead, Year 2, School 4).

In schools where there had been no collaboration across the leadership programmes, staff reflected that Achieve Together had had minimal impact on school improvement. In particular, school leads identified timing issues (such as the collaborative planning meeting taking place too late) as hindering effective implementation and limiting impacts:

> *'I mean it may have worked very well elsewhere but our experience was certainly a little bit back to front and, because of that, the Achieve Together wrapper, so to speak, has had little or no impact'* (in-school lead, Year 2, School 5).

In other cases, staff did not feel the core activities of the programme (such as needs analysis, collaborative planning meeting, and support activities) added sufficient value to impact on outcomes:

> *'I don't think it's had a huge impact as a programme in its own right. I think the idea of trying to encourage collaboration at multiple levels is a good idea … But I think the realities of how they tried to support schools in doing that were probably quite limited'* (in-school lead, Year 2, School 6).

Pupil outcomes

Staff anticipated that the leadership skills developed from participating in individual leadership programmes would ultimately feed into improved pupil attainment, but that this was a long-term goal that would not be apparent immediately:

> *'I mean hopefully the things that I've picked up from it, I'll continue to use over the next couple of years, and then that will hopefully mean that I'm doing my job better which will then have a direct impact on the staff in the team which will in turn have an impact on the results, so I think … could be quite long-term'* (TL Fellows participant, Year 2, School 4).

This view was supported by the project manager who reflected that the two-year timescale of the pilot may have been too short to achieve pupil impacts, particularly because the Achieve Together model of school improvement focused on developing staff capacity, with a view to improving pupil outcomes as

a result. For schools that had successfully implemented collaborative impact initiatives, examples of positive outcomes for pupils included:

- Increased engagement in learning
  In the case of a school that had used its impact initiative to develop and roll out a new approach to attitudes to learning, staff identified positive outcomes for pupil levels of engagement in lessons and there was an expectation that this would ultimately improve outcomes:

  'I think they massively helped engagement actually … [the approach has] enabled the pupils to become a lot more reflective … most of the time now the students are in the habit of just reflecting on each lesson themselves' (Teach First participant, Year 2, School 2).

- Increased independent learning
  Staff identified increased independent learning resulting from an impact initiative aimed at improving attitudes to learning. By supporting pupils to assess their own learning and identify areas where they needed to improve, the expectation was that these skills would ultimately improve attainment outcomes.

Staff also identified positive outcomes for pupils as a result of individual impact initiatives carried out by staff participating in the leadership programmes, although the collaborative approach encouraged by the Achieve Together model did not form part of them. Examples included improved attainment in maths for a peer-tutoring intervention, and improvements in attainment in science for an impact initiative tackling underachievement in this area.

Formative findings

The process evaluation identified the following ways in which Achieve Together could be developed in the future:

- Greater clarity of aims and expectations
  The Achieve Together model relies on schools taking the lead and driving forward collaborative impact initiatives. In light of this, greater clarity of aims and expectations is needed at the outset to structure expectations, foster school engagement and limit delays in implementation.

- Sufficient planning and lead-in time
  To maximise the opportunities for collaboration across the leadership programmes, build in sufficient lead-in time prior to the start of programmes to plan the focus of Achieve Together and communicate this effectively to participating staff:

  'That's the key and really you need to be having that conversation … six months before to say, "Right, you're going onto the programme in such-and-such a time. You've got [this] group of people; your Teach First, your Fellows and your NPQSL. As a school we want that group of people to deliver together … but you're going to have different foci of the same thing". Then it would work, but it needs to be really planned out' (in-school lead, Year 1, School 4).

- Review programme lengths and structures to improve compatibility
  Review the compatibility of the leadership programmes offered within Achieve Together (for example, their length and assessment criteria) to facilitate and foster collaboration.

- More opportunities for collaboration across schools
  School leads reported they would value more opportunities to meet and share learning with other Achieve Together schools. This was an initial aim of the intervention that was removed as it would have affected the strength of the evaluation given that schools would have to be analysed within

their local clusters, not as individual schools. Exploring how this could be embedded into the programme in future should be explored.

Conclusion

Key conclusions
1. Pupils in Achieve Together schools made no additional progress, on average, compared to other children in the trial, as measured by GCSE results. The three padlock security rating means we are moderately confident in this result.
2. Similar results were found for children eligible, and not eligible, for free school meals, and children with higher and lower prior attainment, although these sub-group results are less secure because of the smaller numbers of pupils.
3. Participating schools reported a lack of clarity over the aims of Achieve Together and how the collaborative project should relate to the individual leadership programmes. The differing timelines of the individual programmes also appears to have hampered the collaborative projects.
4. Views on the collaborative project were mixed. Some participants found that the process increased reflective practice and provided impetus for further school improvement. Others felt that it was resource intensive and had minimal additional impact.
5. Some teachers felt Achieve Together would have a positive impact on pupil attainment once new teaching skills and new initiatives were embedded. The EEF data archive allows analysis of future results for these schools to assess whether this is the case.

Interpretation

There is no evidence that Achieve Together had a significant impact on pupil attainment at the end of secondary school, as measured by the primary outcome (capped KS4 points including equivalent qualifications) and all secondary outcomes. The main effects and upper bound of the effects are classified as 'low'. This is true for all sub-groups of pupils considered: pupils eligible, or not eligible, for free school meals, and pupils with higher or lower prior attainment.

The quantitative results are not consistent with previous research undertaken by the three individual partner charities. For example, when compared to comparable schools, Future Leaders report a higher percentage of disadvantaged pupils achieving expected progress in schools where headteachers are involved with their programmes. Teaching Leaders report that when Teaching Fellows are present in a school, a higher percentage of pupils with low prior attainment make expected progress in maths when compared with pupils in eligible (but non-participating) schools. For Teach First, Allen and Allnut (2017) find that the introduction of Teach First results in a school-wide increase in pupil attainment of around 5% of a standard deviation (or one grade in one of the pupil's best eight subjects). Comparing across departments with, and without, a Teach First participant also shows positive effects in the second year after involvement with Teach First, although not statistically significant positive effects for maths and not in the first year of participation (which is similar to our evaluation). The fact that we fail to find a positive impact of these programmes when they are combined clearly runs against this previous evidence.

What could explain the differences compared with these previous studies? First, given that none of these studies are experimental (including our evaluation of Achieve Together) it is possible that one or more of these studies is biased by confounding factors. Second, we note that there was no prior involvement with Teaching Leaders or The Future Leaders Trust prior to involvement with AT, whereas in the case of Teach First some schools already had Teach First placements. In other words, in the case of the former organisations, it was the first time their programmes had been used, whereas in the case of the latter there was an increase in existing involvement. It may be that the effect of moving from little (or no) participation to some participation (as Allen and Allnut suggest) is higher than moving from some to high participation. Third, these programmes were combined with strategic alignment activities which were not universally popular among schools and could have diluted the impact of the individual

programmes rather than increase it. Fourth, we do not know how many of the participants in Achieve Together taught pupils taking GCSEs, whereas the other studies sought to more directly examine the attainment of pupils taught by participants. In general, however, caution is needed here as not only are data and sample sizes insufficient to investigate the source of this difference, but also we are not able to look at the individual programmes in our evaluation.

The evaluation design was a well-matched comparison group rather a randomised controlled trial. This change in design was prompted by low levels of recruitment across two cohorts. The fact that recruitment was challenging is a finding in itself as it may limit the scalability of the programme if schools are unwilling to participate. It also limits the confidence we can have in the effect size for the primary outcome as there may be systematic differences between the intervention schools that signed up for the programme and the set of comparison schools chosen. The likelihood of such systematic differences has been reduced by selecting comparison schools from those that at least expressed an interest in participating and selecting the most comparable schools according to observable characteristics. The balance of observable characteristics between intervention and comparison schools was good, with no marked differences in school-level characteristics.

The Achieve Together model set out to increase alignment and collaboration between individual leadership development training and whole-school priorities. The process evaluation found that in principal the concept of greater collaboration was welcomed by schools, with particular value placed on the benefits of staff working together to share ideas, learn from each other, and reduce workload. In practice, however, the process evaluation found limited evidence of collaboration across the leadership programmes and lessons can be learnt from the initiative, both in terms of feasibility and implementation.

In terms of the feasibility of the model, the process evaluation identified the following lessons:

- Questions were raised over the feasibility of collaboration and alignment across staff. In particular, collaboration was felt to be undermined where staff were (a) not in a position to effectively engage with projects (for example Teach First participants at an early stage of their careers), (b) where impact initiatives did not fit well with staff roles, and (c) where the type of impact initiative chosen was not sufficiently broad in scope to be relevant to staff in different roles and at different stages of their careers.

  Arguably, none of these caveats are insurmountable and with careful planning impact initiatives could be designed to address these concerns.

- The 'backbone support' provided by the initiative—in the form of the needs analysis process, a collaborative planning session, and support activities in the form of webinars and CPD resources—received mixed feedback. Schools were positive about the opportunities for collaboration this support provided and valued some of the planning tools used. However, some questions were raised over the extent to which these activities were sufficient in scope and value to support schools to drive forward a collaborative approach to school improvement. Concerns were also raised over the costs (in terms of staff time and cover costs) required to attend the collaborative planning session.

- The Achieve Together model was designed as a 'self-led' model of school improvement: schools were expected to drive forward the model with support from the Achieve Together team. In some instances it was felt that schools did not have the capacity to take on this role (due, for example, to instability in senior leadership, financial deficits, and so on) which raises questions over the feasibility of the model for schools in the most challenging circumstances.

- Positive impacts of participating in Achieve Together were identified by some schools. These included leadership skill development, greater school improvement impetus, and increases in reflective practice. However, it was felt that the benefits of these impacts would take time to

filter down to improvements in pupil attainment and this may, in part, account for no pupil attainment outcomes identified in the impact evaluation.

In terms of implementation, the process evaluation identified the following lessons:

- The aims of Achieve Together were not well understood by participating schools. Given the 'school-led' nature of the model, this lack of understanding undermined implementation because school staff were unclear of their roles and how the model was intended to work. Clearly articulating roles and expectations at the outset, as well as setting out the nature of the support provided by the Achieve Together team, would improve implementation.

- The timing of Achieve Together project activities undermined successful implementation. Collaborative planning sessions did not take place until the spring term of the first year, by which point it was often too late to align impact initiatives across leadership programmes. Collaboration was more successful in instances where the timing of the support provided by Achieve Together preceded decisions on the impact initiatives that formed a part of the individual leadership programmes.

- The leadership programmes that formed the foundation of the initiative varied in length and assessment criteria. Collaboration across programmes could be enhanced by further consideration being given to the compatibility of the programmes and how they could co-ordinate to maximise opportunities for across-programme collaboration and learning.

Limitations

The intervention took place in 14 schools: this necessarily limits the generalisability of the results to a wide range of schools. The low take-up rate of the programme also suggests that the group of intervention schools may have had distinct requirements from a broader population of schools.

The trial design was quasi-experimental rather than an RCT, which means that there is scope for bias if the well-matched comparison set of schools are systematically different to the schools that signed up to the intervention. This possibility has been limited by choosing the well-matched comparison schools from those that expressed an interest in Achieve Together and using propensity score matching to find the most observable similar group within this.

The outcome measure (GCSEs) was selected as appropriate for a programme aiming for whole-school change, but we do not know how many of the participants were directly involved in teaching the GCSE cohort.

The process evaluation did not evaluate the individual leadership programmes (e.g. Teach First, NPQML, NPQSL, TL Fellows and Future Leaders flagship programme) that formed the foundation of the Achieve Together model as this was beyond the scope of the evaluation. Rather, the focus was on the added value of the Achieve Together model—specifically the collaboration and alignment across the leadership programmes. As a result, no conclusions on the value of the individual leadership programmes can be drawn from the findings of the process evaluation.

Future research and publications

The quantitative and qualitative research from this evaluation suggest that Achieve Together could have positive longer-term impacts on pupil attainment as leadership programmes and collective impact projects become embedded in the school. This should be explored in the future using administrative data to track attainment in the intervention and well-matched comparison group schools. A separate EEF report will analyse the effects of the whole-area Achieve Together programme in Bournemouth.

References

Allen, R. and Allnutt, J. (2017) 'The impact of Teach First on pupil attainment at age 16', *British Educational Research Journal* (forthcoming).

McIntyre, J. and Thomson, P. (2016) 'Poverty, Schooling, and Beginning Teachers Who Make a Difference: A Case Study from England' in *Teacher Education for High Poverty Schools*, Springer International Publishing (pp. 153–170).

Ritchie, J., Lewis, J., McNaughton-Nicholls, C. and Ormston, R. (2013) *Qualitative Research Practice: A Guide for Social Science Students and Researchers*, London: Sage.

The Future Leaders Trust (2015) '2015 Impact Report'. Available at: <http://www.future-leaders.org.uk/documents/160/ImpactReport2015_TzHlcwB.pdf> [Accessed 28 February 2017].

Appendix A: 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 B: Security classification of trial findings

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	Power	Attrition[13]					
5 🔒	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%					
4 🔒	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%			Adjustment for Balance [n/a]		
3 🔒	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%	3				3
2 🔒	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%			Adjustment for threats to internal validity [n/a]		
1 🔒	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	41-50%					
0 🔒	No comparator	MDES > 0.6	over 50%					

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = 3 padlocks
- **Reason for adjustment for balance** (if made): n/a
- **Reason for adjustment for threats to validity** (if made): n/a
- **Final padlock score:** initial score adjusted for balance and internal validity = 3 padlocks

The design is a well-matched comparison group using propensity score matching from a group of schools who expressed interest in the original proposed trial. MDES is 0.19 at randomisation, attrition is approx. 3% for attainment variables in both groups.

---

[13] Attrition should be measured at the pupil level (even for clustered trials) and from the point of randomisation to the point of analysis.

Appendix C: Strategic Alignment Activities

Achieve Together Strategic Alignment Inputs

YEAR 1			YEAR 2		POST ACHIEVE TOGETHER
Scoping	Implementation	Review	Scoping	Implementation	Legacy
✓ Needs analysis interview with Head teacher and lead practitioner (circa one hour) ✓ Analysis of key school strategic documentation including School Improvement Plan, Self-Evaluation Framework, OFSTED Report, target specific action plans and subject audits ✓ Review of three years of retrospective performance data	✓ Regular phone support with lead practitioner (termly or variable according to need/demand) ✓ One whole school collaborative planning training day (circa three hours)	✓ Literature Review of school improvement drivers ✓ Mapping of key school priorities to evidence based drivers of school improvement	✓ Action Research Route Map co-produced with Centre for Use of Research and Evidence in Education (CUREE) ✓ Review of three years of retrospective performance data (termly)	✓ Regular phone support with lead practitioner (termly or variable according to need/demand) ✓ Webinars and newsletter featuring education thought leaders (termly) ✓ AQR Mental Toughness Questionnaire with senior leaders (measuring leadership mindset)	✓ Legacy planning interview with Head teacher and lead practitioner (circa one hour) ✓ Corwin 'I Know My Class' student survey (all KS4 students) ✓ SSAT Embedding Formative Assessment resource pack (subsidy) ✓ Whole Education 'Spirals of Enquiry' (signposting)

Appendix D: Theory of change
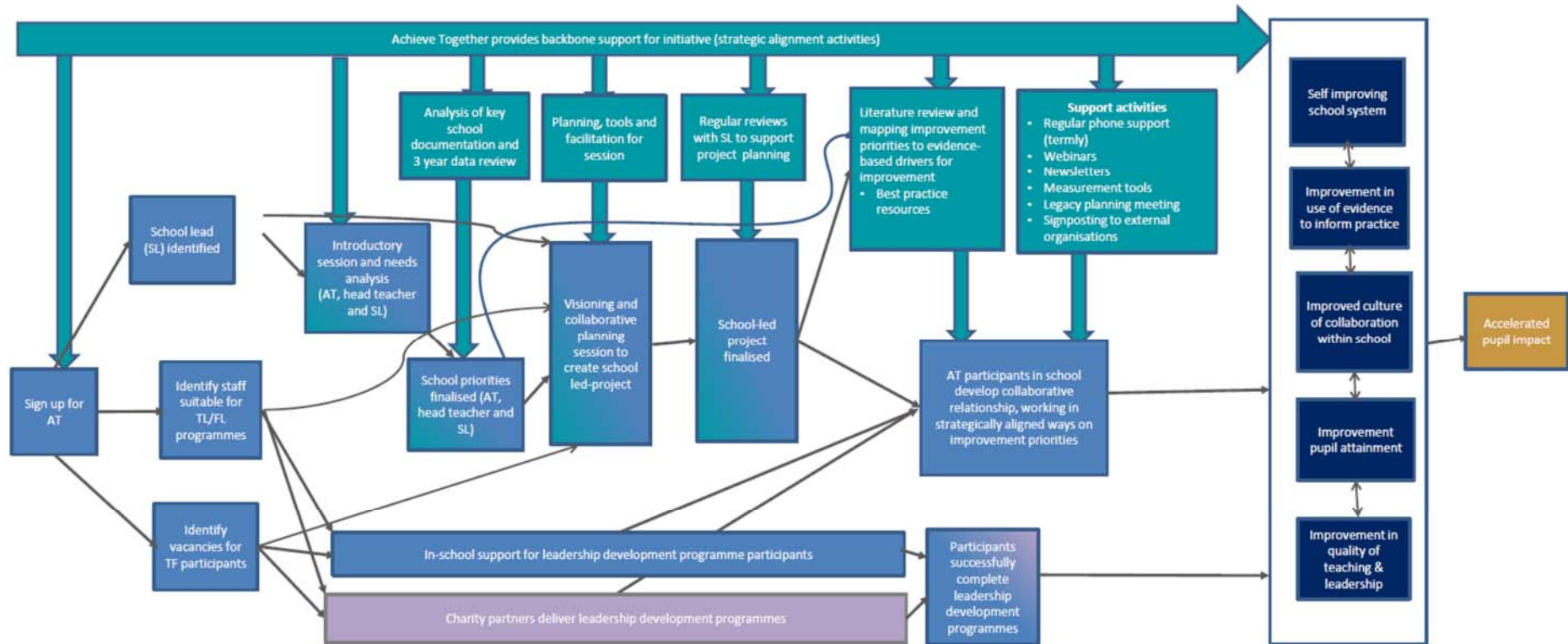
Appendix E: Memorandum of Understanding

Agreement to participate in the Achieve Together Trial

School Name: _____ (the "**School**")

1. Introduction to the Project

Achieve Together is an initiative delivered by three education charities - Teach First, Teaching Leaders and Future Leaders (the "**Charity Partners**") – working together to develop practical ways to attract, retain and develop excellent teachers, middle leaders and head teachers. We believe that a high-quality, co-ordinated investment in teaching and leadership can significantly improve the academic attainment of pupils from low-income communities.

The Achieve Together Trial (the "**Trial**") is funded by the Education Endowment Foundation and consists of a two year intervention aimed at improving educational outcomes through improving the quality of teaching and leadership in participating schools.

There are two main elements to the Trial: a training and development project (as detailed in section 2 below) and an impact project (as detailed at section 3 below). The results of the research will contribute to our understanding of how and to what extent a high quality, whole school coordinated approach to training and developing staff at all levels of a school can support improvements in pupil progress. This evidence is vital: it will be widely disseminated to schools in England to inform their spending decisions, and may determine whether national funding is made available for the Achieve Together model.

2. Overview of the training element of the Trial

The Charity Partners will work with schools in the Trial to create a coordinated approach to training and developing staff, focusing on delivering improvements in pupil progress as follows:

- Future Leaders will develop existing senior members of staff or place senior leaders.
- Teaching Leaders will provide specialist training to promising middle leaders.
- Teach First will place selected new teachers in participating schools.

Achieve Together therefore incorporates the provision of new high-potential teachers and the provision of training and support for existing staff. Each participating school will need to work with all three of the Charity Partners. Ideally schools will have twelve Achieve Together participants within their school during the period of the Trial, spread across at least one each of the Charity Partner's programmes (potentially all five).

3. Overview of the impact project

Alongside this training, schools will undertake an "impact project" (the "**Impact Project**"). The primary focus of each school's Impact Project will be an area identified by the school and an additional focus will be on raising Year 7 pupil attainment. All teachers involved in Achieve Together will contribute towards the Impact Projects.

The success of these Impact Projects will be evaluated through scheduled testing of Year 7 and Year 8 pupils (the "**Evaluation**"), as set out in the evaluation timetable in section 6 below.

4. Timing of the Trial

A full timetable for the Trial is set out at Appendix 2, but the Trial will broadly track the following timescale:

- Forty eight schools will be recruited to The Trial by March 2014.

- In March 2014, forty eight schools will be randomly allocated to one of two groups – the "treatment group" and the "control group" – by the Trial's independent evaluators. This is the only robust way to determine the impact of Achieve Together on pupil progress, which is crucially important to know.
- The Trial will be run in the twenty four schools between September 2014 and July 2016.
- Analysis of data collected in the Evaluation will be provided to schools in June 2015 and June 2016.

5. Overview of Participation in the Trial

As detailed in the section above, schools that are selected to take part in the Trial will be allocated to either the "Treatment Group" or the "Control Group". Participation in both elements of the Trial will mean the following for schools:

Area	Control Group Schools	Treatment Group Schools
Trial Focus	Progress of year 7 students over two years	Progress of year 7 students over two years
Partners	May work with one out of three charities	Will work with all three charities
Timeframe	Will be part of the trial for two years	Will be part of the trial for two years
Testing	Will receive two-year software license free of charge and can be supported to test pupils twice during the trial period	Will receive two-year software licenses free of charge and can be supported to test pupils twice during the trial period
Analysis	Detailed feedback on pupil performance after each of the two tests	Detailed feedback on pupil performance after each of the two tests
Other evaluation	Questionnaire about the trial plus in-depth interviews with some schools in this group	Questionnaire about the trial plus in-depth interviews with some schools in this group
Financial incentive	£2,500 at the end of the trial to be used towards any of the charities' programmes	Subsidies to participate in training programmes
Point of contact	An Achieve Together lead contact for support with admin and organisation	An Achieve Together lead contact for support with admin and organisation
Wider moral purpose	Will join the growing number of effective schools that are involving themselves in national educational research to benefit their own students and the sector as a whole	Will join the growing number of effective schools that are involving themselves in national educational research to benefit their own students and the sector as a whole

Schools that are allocated to the Control Group and already access the programme(s) of one of the Charity Partners can continue to do so.

Schools allocated to the Treatment Group must commit to purchase at least one programme from each of the Charity Partners. Acceptance of this document represents an intention to purchase these programmes should a school end up as part of the Treatment Group. Details of the exact programmes a school will purchase are detailed below.

Please indicate below the number of participants you will be taking on as part of the Achieve Together initiative.

Programme	# of participants
Teach First	
Teaching Leaders – NPQML	
Teaching Leaders – Fellows	
Future Leaders – Flagship	
Future Leaders – NPQSL	

6. Structure of the Evaluation

The Evaluation is an integral part of the intervention and is being conducted independently by the Institute for Fiscal Studies ("**IFS**") and the National Centre for Social Research ("**NatCen**").

The full outline programme Evaluation timetable is included in Appendix 2, but a summary of the key elements of the Evaluation is listed below.

The outcome measures for the programme are:

- Attainment in Year 7 (Hodder access reading and access numeracy tests)
- Attainment in Year 8 (Hodder access reading and access numeracy tests)
- Attainment at Key Stage 4 (capped points score; number of qualifications; subject and type of qualifications)
- Absence (number of persistent and total absences)

The outcomes are listed in order of relevance to the pilot, where the impact on attainment is the **primary outcome** and the impact on absences is the **secondary outcome**.

- **Primary outcome (impact on attainment):** The quantitative evaluation will be led by the IFS and will focus on pupil attainment. It will consist of a formal assessment (using Hodder access reading and access numeracy tests) for pupils in year 7 and year 8, in both May 2015 and 2016, and will compare these results with pupils' previous attainment at the end of Key Stage 2. Analysis of this data will be provided to schools in June 2015 and 2016. Data from the June 2015 and 2016 GCSEs will also form part of the outcome indicators.

- **Secondary outcome (impact on absences):** As a secondary outcome the IFS will use pupil absence levels. This data will be collected via the National Pupil Database ("**NPD**") in May 2015 and May 2016.

NatCen is responsible for the process evaluation. The process evaluation methods will be:

- a survey of all schools participating in the pilot; and
- depth interviews with 10 schools participating in the pilot, including the collection of demographic data.

Each school selected for the depth interviews will be spoken to in each year of the pilot (around January). Within each school NatCen will talk to a range of staff participating, including one of each of the following the Head teacher, Achieve Together participants and the In-School Lead.

7. Responsibilities

Below are details of individual responsibilities necessary to ensure the successful delivery of Achieve Together. Key milestones for the programme can be found in Appendix 1.

Achieve Together will:

- Set out the requirements of the programme through information materials and follow up with schools;
- Maintain a core programme team, including dedicated partnership officers, to support schools through the Full Trial;
- Be the first point of contact for queries from schools in relation to the Full Trial;
- Support schools to ensure the successful placing of trained Teach First participants;
- Support schools with the process of putting staff forward for Teaching Leaders and Future Leaders programmes;
- Support schools with the development, implementation and review of their impact projects;
- Share information about the programme across the participating schools and facilitate and deliver opportunities for collaboration between interested schools;
- Support schools with the implementation of the testing elements of the evaluation;

- Support schools to work with the IFS and NatCen to maximise the value of the process and quantitative evaluation elements of the Trial;
- Ensure control schools are able to access their subsidy as appropriate at the culmination of the programme.

The School will consent to participate fully in the Achieve Together programme, the meaning of which will be determined by its status as either a "treatment school" or a "control school" (please refer to section 5 above for further detail on these). Schools will:

- Identify staff and put them forward for the Teaching Leaders and Future Leaders programmes (dependent on their treatment or control group status);
- Purchase from the Charity Partners the requisite training and support packages (dependent on their treatment or control group status);
- Support all Achieve Together participants to engage effectively with the programme;
- Develop and deliver impact projects aligned with the Achieve Together outcomes;
- Fully participate in the evaluation;
- Allow time and provide sufficient staff resources to implement for testing year 7 and year 8 pupils in reading and numeracy in May 2015 and May 2016;
- Provide consent for the IFS to access NPD records for pupil data necessary for the evaluation.

To enable this to happen we request that Head teachers:

- Champion the Achieve Together initiative in their school;
- Ensure that the Achieve Together impact project aligns with the school's development priorities;
- Act as the main point of contact for Achieve Together until such time as they are able to;
- Appoint, from within the Achieve Together cohort, an In-School Lead who will be the main point of contact for Achieve Together during the Full Trial.

The IFS will:

- Conduct the random allocation;
- Collect and analyse data from the NPD for the project;
- Write the draft report;
- Disseminate research findings.

NatCen will:

- Support the implementation of the Year 7 and Year 8 reading and numeracy tests;
- Provide head teachers with analysis of data from testing;
- Lead the questionnaire process in schools;
- Write the draft report;
- Disseminate research findings.

8. General

Achieve Together is aware that Schools may wish to use the Achieve Together name or logo for the purposes of championing the Trial in their school, and the School is permitted to do so. The School agrees, however, to obtain prior written consent from Achieve Together before using the name or logo of Achieve Together in any external facing materials or for any other purpose not related to the Trial.

Achieve Together is also fully committed to ensuring that any confidential information or data it receives during the Trial is only used for the purposes of the Trial. The School agrees that it will also ensure that the confidential information of Achieve Together is only used for the purposes of the Trial and will not disclose or pass to any other party the confidential information of Achieve Together.

The Charity Partners are fully committed to working closely with Schools to improve educational outcomes through improving the quality of teaching and leadership. However, in the event of a dispute or any concerns being raised by either party regarding the Trial, Achieve Together and the School will attempt to settle that dispute by amicable negotiation. In such negotiations, the Achieve Together Manager will represent Achieve Together and the School's Head teacher will represent the School. Where a dispute arises between a Charity Partner and the School, this will be resolved directly between the parties to the dispute in accordance with the separate agreements entered into between the relevant Charity Partner and the School.

If the parties are unable to settle any dispute by amicable negotiation within 30 days of the first meeting between these negotiators, either party may refer the dispute to the Centre for Effective Dispute Resolution ("**CEDR**") in accordance with CEDR's model mediation procedure.

Both parties are fully committed to the Trial and will use their best endeavours to reach an amicable solution to any dispute. In the event that it is not possible to settle a dispute between Achieve Together and the School, within 30 days of discussions commencing, the parties may explore bringing their relationship, or part of their relationship, and the Trial to an end. The School acknowledges and agrees that in these circumstances, it may no longer be eligible to receive any financial incentives offered as part of the Trial (as detailed at section 5), including subsidies for treatment schools, and Achieve Together will have full discretion to determine whether or not the School will continue to receive this. Any use of the programme of one of the Charity Partners may continue notwithstanding the early termination of the Trial, and in that event, the School agrees that it will continue to provide full support to any participants of a Charity Partner's programme.

The parties intend this section 8 to be legally binding, and governed by the laws of England and Wales and subject to the exclusive jurisdiction of the courts of England and Wales. The remainder of this agreement are not intended to be legally binding.

9. Next steps

Immediately after the random allocation has taken place (end of March 2014) Achieve Together will contact all schools to confirm the outcomes and discuss next steps.

For treatment schools this will include linking schools, where necessary, into each of the Charity Partner organisations to enable them to move forward with their individual programmes. It will also include initial discussions around the potential shape of Impact Projects schools may want to develop for September 2014.

10. Agreement

We commit to fully participate in the Achieve Together Trial on the terms detailed above and will support the delivery of the Intervention as required.

Signatures

On behalf of Achieve Together

Achieve Together Partnership Lead [NAME]: _____

Date: _____

Signature: _____

On behalf of the School named on page 1
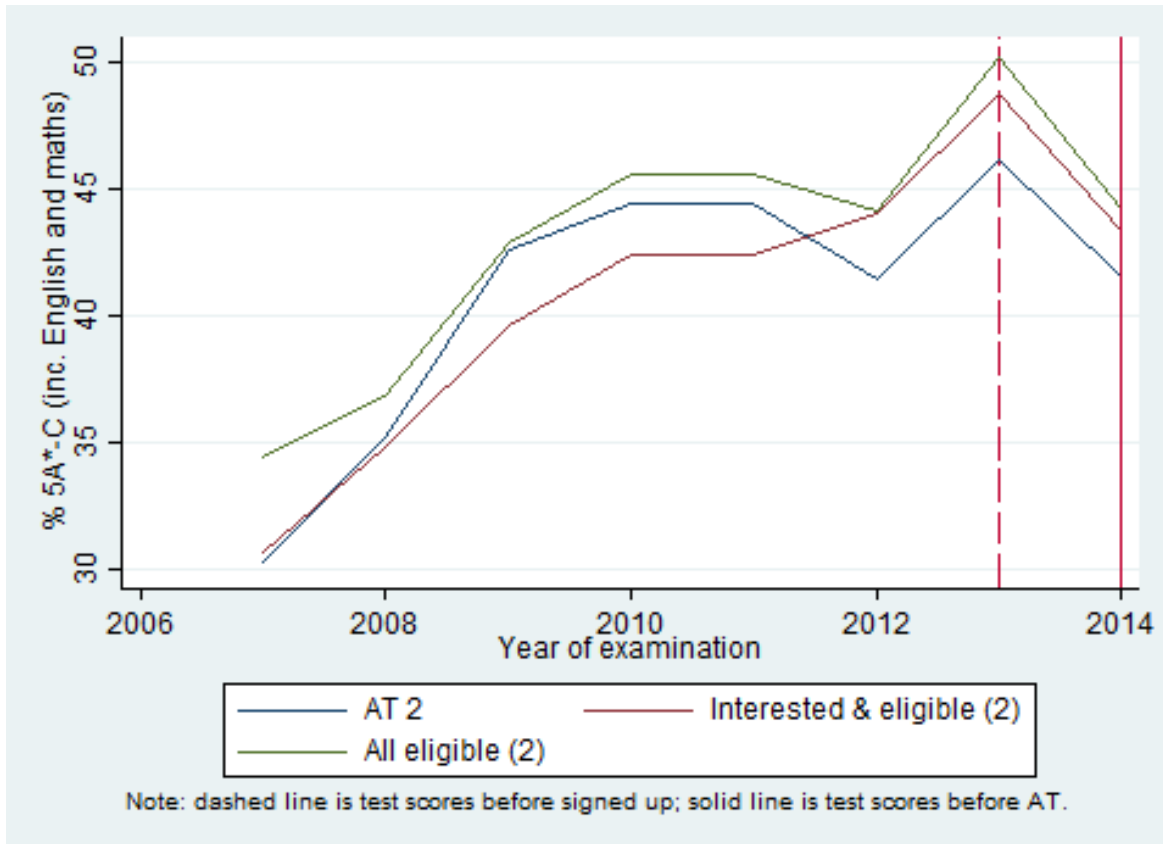
Head teacher [NAME]: _____

Date: _____

Signature: _____

Once signed please return to Chris Gurney, Achieve Together Manager, Teach First, 4 More London Riverside, London, SE1 2AU.

Appendix F: Additional analyses

Appendix Figure 1: Long-run trends in pupil attainment before Achieve Together

"AT 2" represents schools who participated in the second cohort of Achieve Together, "Interested and Eligible (2)" represents schools who were interested in participating but not actually sign up, "All Eligible (2)" represents all schools who were eligible for the second cohort of Achieve Together.

Appendix Table 1: Sensitivity analysis (end of year two of Achieve Together imposing a common sample for all outcomes)

Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Effect size (95% CI)	p-value
KS4 points score (2016)	2328 (51)	-0.192 (-0.228;-0.157)	5974 (157)	-0.173 (-0.190;-0.156)	53 (14;39)	0.016 (-0.061;0.093)	0.682
KS4 points score (excluding equivalents) (2016)	2328 (51)	-0.330 (-0.368;-0.292)	5974 (157)	-0.347 (-0.366;-0.329)	53 (14;39)	0.002 (-0.103;0.106)	0.976
Teach First impact goal measure (2016)	2328 (51)	-0.315 (-0.353;-0.277)	5974 (157)	-0.325 (-0.343;-0.306)	53 (14;39)	0.001 (-0.102;0.104)	0.983
Overall absence (2016)	2328 (51)	0.028 (-0.013;0.069)	5974 (157)	0.019 (-0.001;0.039)	53 (14;39)	-0.024 (-0.072;0.024)	0.315

Stars for statistical significance: $p <= 0.10$ (*) $p <= 0.05$ (**) $p <= 0.01$ (***)

A common sample has been imposed for all four outcome variables –i.e. observations missing data on one or more of the four outcome variables have been excluded from the sample.

Standard errors clustered at school level.

Propensity score weights are applied to the control group.

Outcomes (aside from absence) are standardised according to the national distribution of scores to have a mean of zero and standard deviation of one.

Teach First impact goal measure represents the capped GCSE point score excluding equivalents plus English and maths bonuses.

Overall absence is measured in number of unauthorised absences over the course of the school year and is presented in standardised terms within the estimation sample.

Pupil-level controls are: season of birth, gender, ever FSM, IDACI score, binary indicator for missing IDACI score, KS2 quartile, binary indicator for missing KS2 quartile. KS2 quartile is based on teacher assessment where test scores are missing due to KS2 boycott in 2010.

School-level controls (2012/2013) are: % achieving 5A*-C GCSE and equivalents including English and maths, total average (capped) point score per pupil (GCSEs only), value added measure based on the best 8 GCSE and equivalent results, % teachers with Qualified Teacher Status, average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching', latest Ofsted rating for 'leadership and management' at least 'good', percentage of pupils eligible for free school meals, percentage of pupils with English as an additional language , percentage of pupils with SEN statement or on School Action Plus.

Appendix Table 2: Sensitivity analysis (including school-level random effects)

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Effect size (95% CI)	p-value
KS4 points score (2016)	2353 (26)	-0.202 (-0.237;-0.166)	6045 (157)	-0.181 (-0.198;-0.164)	77 (14;40)	0.012 (-0.068;0.093)	0.754
KS4 points score (excluding equivalents) (2016)	2353 (26)	-0.341 (-0.379;-0.302)	6045 (157)	-0.354 (-0.373;-0.336)	77 (14;40)	0.012 (-0.094;0.118)	0.814
Teach First impact goal measure (2016)	2353 (26)	-0.326 (-0.364;-0.287)	6045 (157)	-0.332 (-0.351;-0.313)	77 (14;40)	0.012 (-0.089;0.112)	0.81
Overall absence (2016)	2328 (51)	0.028 (-0.013;0.069)	5974 (228)	0.019 (-0.001;0.039)	76 (14;39)	-0.033 (-0.087;0.021)	0.208

Stars for statistical significance: p <= 0.10 (*) p <= 0.05 (**) p <= 0.01 (***)

This specification allows for school-level random effects.

Standard errors clustered at school level.

Propensity score weights are applied to the control group.

Teach First impact goal measure represents the capped GCSE point score excluding equivalents plus English and maths bonuses.

Overall absence is measured in number of unauthorised absences over the course of the school year and is presented in standardised terms within the estimation sample.

Outcomes (aside from absence) are standardised according to the national distribution of scores to have a mean of zero and standard deviation of one.

Pupil-level controls are: season of birth, gender, ever FSM, IDACI score, binary indicator for missing IDACI score, KS2 quartile, binary indicator for missing KS2 quartile. KS2 quartile is based on teacher assessment where test scores are missing due to KS2 boycott in 2010.

School-level controls (2012/2013) are: % achieving 5A*-C GCSE and equivalents including English and maths, total average (capped) point score per pupil (GCSEs only), value added measure based on the best 8 GCSE and equivalent results, % teachers with Qualified Teacher Status, average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching', latest Ofsted rating for 'leadership and management' at least 'good', percentage of pupils eligible for free school meals, percentage of pupils with English as an additional language , percentage of pupils with SEN statement or on School Action Plus.

Appendix Table 3: Secondary analysis (year one estimates for Achieve Together cohort 1)

| Outcome | Raw means | | | | Effect size | | |
| | Intervention group | | Control group | | | | |
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Effect size (95% CI)	p-value
KS4 points score (2015) [interested schools as control]	2239 (0)	-0.084 (-0.118;-0.049)	3529 (57)	-0.195 (-0.223;-0.168)	35 (14;21)	-0.004 (-0.078;0.071)	0.923
KS4 points score (2015) [cohort 2 schools as control]	2239 (0)	-0.084 (-0.118;-0.049)	1287 (15)	-0.242 (-0.278;-0.207)	22 (14;8)	0.027 (-0.089;0.143)	0.629

Stars for statistical significance: p <= 0.10 (*) p <= 0.05 (**) p <= 0.01 (***)

These are estimates on outcomes after the first year of Achieve Together for cohort 1 and the number of schools differ from cohort 2 analysis as a direct result. Row (1) uses schools interested in Achieve Together during recruitment for cohort 1. Row (2) uses cohort (2) schools as a control group.

The controls for each specification are as described above.

Standard errors clustered at school level.

Propensity score weights are applied to the control group.

Outcomes (aside from absence) are standardised according to the national distribution of scores to have a mean of zero and standard deviation of one.

Pupil-level controls are: season of birth, gender, ever FSM, IDACI score, binary indicator for missing IDACI score, KS2 quartile, binary indicator for missing KS2 quartile. KS2 quartile is based on teacher assessment where test scores are missing due to KS2 boycott in 2010.

School-level controls (2012/2013) are: % achieving 5A*-C GCSE and equivalents including English and maths, total average (capped) point score per pupil (GCSEs only), value added measure based on the best 8 GCSE and equivalent results, % teachers with Qualified Teacher Status, average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching', latest Ofsted rating for 'leadership and management' at least 'good', percentage of pupils eligible for free school meals, percentage of pupils with English as an additional language , percentage of pupils with SEN statement or on School Action Plus.

Appendix Table 4: Sensitivity analysis (using teacher assessments of KS2 for all pupils)

| | Raw means | | | | Effect size | | |
| | Intervention group | | Control group | | | | |
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Effect size (95% CI)	p-value
KS4 points score (2015)	2418 (33)	-0.222 (-0.256;-0.189)	6372 (144)	-0.158 (-0.174;-0.141)	54 (14;40)	-0.068 (-0.147;0.010)	0.082*
KS4 points score (excluding equivalents) (2015)	2418 (33)	-0.349 (-0.386;-0.312)	6372 (144)	-0.338 (-0.356;-0.320)	54 (14;40)	-0.069 (-0.156;0.018)	0.113
Teach First impact goal measure (2015)	2418 (33)	-0.326 (-0.363;-0.290)	6372 (144)	-0.313 (-0.332;-0.295)	54 (14;40)	-0.068 (-0.145;0.009)	0.077*
Overall absence (2015)	2394 (57)	-0.047 (-0.081;-0.013)	6281 (235)	-0.046 (-0.064;-0.029)	54 (14;40)	0.004 (-0.043;0.050)	0.88

Stars for statistical significance: p <= 0.10 (*) p <= 0.05 (**) p <= 0.01 (***)

This analysis uses teacher assessments of KS2 for all pupils. This is performed as a sensitivity analysis because a KS2 boycott in 2010 means that KS2 scores are missing for around a third of pupils – for this cohort KS2 test scores are imputed using KS2 teacher assessments in the main analysis.

Standard errors clustered at school level.

Propensity score weights are applied to the control group.

Outcomes (aside from absence) are standardised according to the national distribution of scores to have a mean of zero and standard deviation of one.

Pupil-level controls are: season of birth, gender, ever FSM, IDACI score, binary indicator for missing IDACI score, KS2 quartile, binary indicator for missing KS2 quartile. KS2 quartile is based on teacher assessment where test scores are missing due to KS2 boycott in 2010.

School-level controls (2012/2013) are: % achieving 5A*-C GCSE and equivalents including English and maths, total average (capped) point score per pupil (GCSEs only), value added measure based on the best 8 GCSE and equivalent results, % teachers with Qualified Teacher Status, average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching', latest Ofsted rating for 'leadership and management' at least 'good', percentage of pupils eligible for free school meals, percentage of pupils with English as an additional language , percentage of pupils with SEN statement or on School Action Plus.