Achieve Together

Future Leaders, Teach First, Teaching Leaders
Institute for Fiscal Studies (IFS)

Evaluation Summary	
Age range	11-16
Number of pupils	c. 12,750
Number of schools	15
Design	Well-matched comparable
Primary Outcome	KS4 GCSE and equivalents

BACKGROUND

Significance

It is widely recognised that the quality of teaching a pupil receives has a significant impact on their achievement in school and in later life. The Department for Education has committed to improving the quality of the teacher workforce, but there is currently little evidence on the most effective way of raising the standard of teaching in England's schools. The question of how to improve aspects of teaching and leadership in areas of high deprivation is of particular importance.

Achieve Together is designed to improve the quality of teaching and management in deprived schools in England. The evaluation will provide evidence on whether an intensive investment in the human capital of teachers in a school leads to an improvement in pupils' academic performance.

Intervention

Achieve Together is an initiative delivered by three education charities - Teach First, Teaching Leaders and Future Leaders – working together to develop practical ways to attract, retain and develop excellent teachers, middle leaders and head teachers. The ultimate aim is to raise standards in schools in low-income communities.

Achieve Together will have a unified focus on improving pupil attainment at all levels of teaching and leadership in a school. Future Leaders will develop existing senior members of staff or place senior leaders. Teaching Leaders will provide specialist training to promising middle leaders and Teach First will place selected new teachers in participating schools. Achieve Together therefore incorporates the provision of new high-potential teachers as demand requires, and training for existing staff.

In line with the original aim of Achieve Together, the intervention focuses on developing the quality of leadership in schools and fostering collaborative responses to whole-school improvement priorities to deliver improvements in pupils' academic performance. The ultimate aim is to develop our understanding of how collaborative working across the leadership tiers of a school can raise academic standards for pupils from low-income communities.

The Achieve Together team will identify best practice in collaborative working and build a model which they will support schools to implement in their setting. These ways of working will be focused around a Collective Impact Project, focused on a key area of school improvement linked into improving outcomes at KS4

Achieve Together will run for two years in the selected schools. Over the two year period each participating school will be expected to work with all 3 partner charities. Ideally schools will be expected to have at least 12 participants, although schools will not be excluded from the pilot if they fail to reach this target.

A cluster of schools in Bournemouth will pilot an area-based approach which includes co-ordination across schools and the involvement of external agencies in addition to the core Achieve Together pilot. The Bournemouth pilot will not be quantitatively evaluated, and therefore is not discussed further below (for more details, see http://educationendowmentfoundation.org.uk/uploads/pdf/Round_2_-_Achieve_Together_Bournemouth.pdf).

RESEARCH PLAN

Research questions

- What is the impact of Achieve Together on average school-level attainment (measured by KS4 attainment)?
- What is the impact of Achieve Together on average attainment (measured by KS4 attainment) of those with low prior attainment and those identified as being eligible for free school meals?
- What is the impact of Achieve Together on the number of persistent absentees and the overall absence rate at the school?

Design

The design of this evaluation has changed in response to the number of schools recruited to the Achieve Together Full Trial. The initial design was a randomised controlled trial, where at least 48 schools would be recruited and 24 schools would be randomly assigned to the treatment group. However, the number of schools eventually recruited to the pilot prohibits a randomised controlled trial design.

Recruitment to the Full Trial 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; complexity of the offer.

Recruitment was low for a number of reasons: uncertainty about staff numbers available to participate; uncertainty about the school budget; the relatively high cost of involvement; the risk of being allocated to the control group. After two attempts to recruit sufficient numbers of schools for the RCT design, there are two cohorts of Achieve Together schools (that were all eventually involved in Achieve Together):

1. Cohort 1: 13 schools recruited to begin Achieve Together in September 2013
2. Cohort 2: 15 schools recruited to begin Achieve Together in September 2014

Schools in Cohort 1 and Cohort 2 were both recruited to the randomised controlled trial and agreed to the conditions this design imposed. There were minor differences to the information given to schools in Cohort 1 and Cohort 2, however:

1. Cohort 1 signed up to a three year intervention, rather than a two year intervention for Cohort 2
2. Cohort 1 had stricter requirements on their behaviour if allocated to a control group in that they were not allowed to participate in any partner programme (whereas Cohort 2 were able to continue to work with one partner if a control school)
3. Cohort 2 were offered more generous and targeted subsidies for some of the programmes. 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
4. Emphasis was placed on the attainment of year 7 and year 8 for Cohort 2
5. Cohort 2 were offered universal testing of year 7 and year 8 pupils

These differences may make Cohort 1 and Cohort 2 schools incomparable in ways that are observable and unobservable to the researcher. For example, schools that signed up to Cohort 2 might have lower performance in year 7 and year 8, relative to KS4. However, the core features of Achieve Together model when schools signed up to the trial remained the same. We therefore hypothesise that schools that form Cohort 2 are largely similar to schools that form Cohort 1 in observable ways (such as prior attainment and teacher vacancies) and less observable ways (such as the school's expected financial position and the head teacher's motivation and belief in the importance of staff development).

The main aim of the quantitative evaluation is to estimate the impact of Achieve Together for Cohort 2, after the first and second year of the programme's introduction. The estimation approach will be through a well-matched comparable group, using two approaches:

1. Choosing a well-matched comparable group using observable information about schools from various sources (Edubase, the School Workforce Census, Ofsted, and the National Pupil Database) that is likely to influence both the decision to participate in the trial and the growth in pupil outcomes over time: prior attainment, the teacher vacancy rate, the teacher turnover, the proportion of pupils eligible for free school meals, the proportion of pupils with English as an additional language, Ofsted ratings.
2. Choosing a well-matched comparable group from the list of schools that showed strong interest in the Achieve Together pilot but did not sign up (perhaps because of the randomisation involved in the initial trial), and a similar set of covariates within this.

These approaches to find a suitable well-matched comparable group are likely to work reasonably well. The comparison of the impact estimates from approach one and two will inform whether schools that sign up for randomised controlled trials are systematically different to schools that do not (but look similar in many other ways). This will in part inform the external validity of similar EEF trials.

These estimates will be biased if schools that were willing to sign up to the Achieve Together pilot have a systematically higher or lower growth in attainment than the well-matched comparable group for reasons that are not observable to us (for example the expected financial position of the school). We will assess the likely direction and magnitude of this bias by using a third approach, which is possible for the impact estimate for the first year of Achieve Together only, but will give the most reliable impact estimate:

3. Using Cohort 2 Achieve Together schools as a comparable group for Cohort 1 Achieve Together schools. These two sets of schools should be similar in observable and unobservable characteristics, given the similarity of the scope and nature of the Achieve Together pilot they signed up to. In the first year of Achieve Together for Cohort 1 schools, the only difference between Cohort 1 and Cohort 2 should be that Achieve Together has begun in Cohort 1 schools but has not yet begun in Cohort 2 schools.

Unfortunately this approach can't be replicated to estimate the longer term impact of Achieve Together for Cohort 1 schools or the impact of Achieve Together for Cohort 2 schools. Instead, this approach will be used to inform whether the first and/or second approaches produce reliable impact estimates. This will be done as follows:

- Estimate the impact of Achieve Together in the first year using approach 1
- Estimate the impact of Achieve Together in the first year using approach 2
- Estimate the impact of Achieve Together in the first year using approach 3
- Compare the impact estimate of Achieve Together in the first year from approach 3 with those from approach 1 and 2.
- If approach 1 and/or approach 2 give a similar impact estimate to approach 3, we can apply the most appropriate method to future years of Achieve Together with confidence.
- If neither approach gives a similar impact estimate to approach 3, we will not be able to provide a causal estimate of the impact of Achieve Together for Cohort 2 schools.

The impact estimate will be compared with the overall costs of the pilot, which will be collected from the Achieve Together project team.

Participants

The main programme will be piloted in 17 schools between September 2014 and July 2016. 15 of these schools signed up to the Achieve Together Full Trial with full knowledge of the original randomised controlled trial design. These 15 schools will be the focus of our main analysis, but the results will be compared to those including the 17 schools.

Treatment and control schools will be spread geographically, but will exclude London given the high engagement many schools in London currently have with either Teach First, Teaching Leaders and Future Leaders. Treatment and control schools must meet the eligibility criteria used by Teach First, Teaching Leaders and Future Leaders: at least 50% of pupils must come from the lowest 30% of the IDACI (Income Deprivation Affecting Children Index); or at least 50% of pupils are eligible for free school meals, or performance at KS4 is below the national average for schools where the proportion of pupils eligible for free school meals is between 25% and 50%.

Outcome Measures

- 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.

Sample size calculations

Table 1 below shows the total sample size required for 80% test power, at a 5% level of significance, depending on the estimated effect size (measured in standard deviations, SDs) and the within-cluster correlation in test scores. We have assumed that there are 170 pupils per year group, per school.

Given the number of schools that will receive the treatment, we have calculated the minimum effect size (in standard deviations) that could be detected under various assumptions of the correlation in outcomes between pupils at the same school (the intra cluster correlation of outcomes).

Model 1 reports the minimum detectable effect size when the variance of the outcome unexplained by attributes of the pupils (including prior attainment) is 60%. Model 2 reports a less optimistic scenario (70% unexplained), whilst Model 3 is more optimistic (50% unexplained)

Table 1 Power calculations

	Intra class correlation					
	0	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**
Minimum detectable Effect Size (Model 1)	0.061	0.187	0.257	0.312	0.359	0.400
Minimum detectable Effect Size (Model 2)	0.066	0.202	0.278	0.337	0.387	0.432
Minimum detectable Effect Size (Model 3)	0.055	0.171	0.235	0.285	0.327	0.365

Note: these calculations represent the effect size that will be possible to detect using a two-sided hypothesis test with significance level of 5%, and with power against an alternative hypothesis of 80%. We have assumed 170 pupils per school for 15 treatment and 15 control schools.

With a realistic level of intra class correlation (around 0.15), we would be able to detect an effect size of around 0.3 standard deviations. This is large relative to existing estimates; the relationship between GCSE attainment and the introduction of a Teach First teacher is around 0.1, for example (Muijs et al, 2010)[1]. Note that if the number of control schools in increased (which would be possible using the well-matched comparable group approach) the minimum detectable effect size is slightly smaller (around 13% smaller) but still large relative to existing estimates.

The sample of schools to be used in the Bournemouth cluster has already been determined. The sample size is not sufficient to estimate the impact quantitatively, but we will provide descriptive statistics and qualitative evidence on the change in academic standards in the primary and secondary schools over time (described in a separate evaluation protocol here http://educationendowmentfoundation.org.uk/uploads/pdf/Round_2_-_Achieve_Together_Bournemouth.pdf).

Analysis plan

The Institute for Fiscal Studies (IFS) is responsible for the quantitative impact evaluation of the Achieve Together pilot.

Evaluating the impact of the pilot on pupil outcomes will include the cost-effective and robust use of administrative data to assess the impact on educational and behavioural outcomes at age 16 (at the end of Key Stage 4). The evaluation will focus on Achieve Together Cohort 2, but Cohort 1 will be used to assess the validity of the well-matched comparable group evaluation design.

The first outcome of interest for the quantitative evaluation will be the attainment of pupils assessed at age 16 (at the end of Key Stage 4) after the first and second year of the programme, measured using the National Pupil Database (NPD). This approach has three main advantages: first, this analysis will be robust to attrition from the pilot, as the administrative data is based on a pupil level census that is a statutory requirement for all state schools in England; second, it will be cost effective, given the expertise of the IFS in the use of this data; third, there is no burden placed on schools through requiring additional testing of pupils.

[1] Muijis et al (2010) “Maximum Impact Evaluation; The impact of Teach First teachers in schools”

The impact of the programme on the number of persistent absentees and the overall absence rate will be estimated from national administrative data. This will provide a limited indication of the attitude to school and behaviour of pupils in the pilot and control schools.

Given the relatively small number of schools in the treatment and control groups, it is unlikely that sub-group analysis by school would reveal significant differences. However, it will be possible to define the achievement results for different groups within individual schools (e.g. we will examine whether there are larger effects for pupils eligible for FSM and pupils with low previous attainment).

As discussed in the "Design" section of this Evaluation Protocol, the likely bias (if any) of these impact estimates using well-matched comparable groups for Cohort 2 will be assessed by computing the impact estimates for Cohort 1 in the first year of Achieve Together using the same methods, and comparing these estimates to the benchmark comparable group for Cohort 1: Cohort 2 Achieve Together schools. As discussed above, Cohort 2 Achieve Together schools are the best comparable group for Cohort 1 Achieve Together schools as these schools had the same motivation for signing up to the Achieve Together pilot. This means that observable and less observable attributes of the schools in the two cohorts are likely to be very similar.

Process evaluation methods

The Achieve Together programme aims to achieve whole school change by facilitating a unified focus on pupil attainment and investing in the development and retention of staff. Therefore, we propose a case study approach to gather feedback from the full range of staff involved taking part to understand how implementation and delivery of the intervention is working over the first two years of the programme.

We propose conducting case study visits with five of the fifteen schools participating in Cohort 2 of the programme which will ensure we capture the full range and diversity of schools taking part. Schools will be purposively sampled to achieve diversity on school level characteristics like size; level of deprivation; and prior attainment.

At the outset Achieve Together will develop a logic model for the intervention that sets out how the outcomes the programme is seeking to achieve and how programme delivery. This will then be used to inform the process evaluation including the topic guides used for data collection.

The following table provides an overview of data collection across the two years which is discussed in further detail below:

Data collection overview

	Year 1: phase 1	Year 1: phase 2	Year 2	Total
Achieve Together project manager	1	1	1	3
Achieve Together school leads	5	5	5	15
Achieve Together teachers	0	15	15	30
Overall total	***6***	***21***	***21***	***48***

Case study visits in Year 1 (September 2014-July 2015)

e propose collecting case study data in two phases in the first year of implementation:

- *Phase 1: Telephone interviews with Achieve Together leads within the five case study schools (November 2014):* To gather early feedback on expectations for the programme, motivations for joining Achieve Together and implementation of the programme, we will conduct telephone interviews with the five Achieve Together leads in the case study schools in the Autumn term 2014. This will also allow the research team to make contact and engage schools in the evaluation process, helping to ensure buy-in for the evaluation going forward. The interviews will be light-touch, lasting no longer than 45 minutes.

  A telephone interview will also be conducted with the Achieve Together project manager to gather their feedback on the implementation of the programme and the nature of any facilitators /challenges to programme implementation.

- *Phase 2: Case study visits (Summer 2015):* The light-touch telephone interviews in Phase 1 will be followed-up by case study visits at the end of Year 1 in all five case study schools. Case study visits will involve:
  - Depth interviews with Achieve Together Lead (approx 45mins)
  - Depth interviews (or mini-groups) with 2-3 other members of staff involved in Teach First, Future Leaders, Teaching Leaders and participating in the programme.

  Interviews will explore with participants:
  - Their experiences of the Achieve Together programme in Year 1
  - Nature of project activities
  - Facilitators / barriers to impact
  - Expectations for the coming year

  A follow-up telephone interview with the Achieve Together mainstage project manager will also be conducted to gather any feedback from implementation and progress in year 1. In total twenty-one interviews will be conducted in phase 2.

Case study visits in Year 2 (September 2015-July 2016)

A final round of case study visits will be conducted in the Summer 2016 to gather feedback and reflections on the programme as it reaches its end. Case study visits will mirror those that took place in the Summer 2015, although individual staff participating may be different depending on availability and staff changes.

A final telephone interview will also be conducted with the Achieve Together Project Manager to gather final reflections and views. A total of twenty-one interviews will be conducted in Year 2.

All interviews will be audio recorded and transcribed to facilitate comprehensive and systematic analysis. Data will be managed in NVivo and analysed using a Framework approach that facilitates descriptive and explanatory analysis.

Process evaluation outputs

The process evaluation will feed into the final report that will be drafted in January 2017. The findings will focus on the implementation of Achieve Together, staff reflections on the programme, facilitators and barriers to impact and key learning and recommendations for programme roll-out.

PERSONNEL

Ellen Greaves	Institute for Fiscal Studies	Overall evaluation manager and head of

		impact evaluation
Luke Sibieta	Institute for Fiscal Studies	Expert advice on the impact evaluation
Research economist	Institute for Fiscal Studies	Working on the impact evaluation
Cheryl Lloyd	NatCen Social Research	Senior oversight for process evaluation
Meg Callanan	NatCen Social Research	Project manager for process evaluation
Hashim Ahmed	NatCen Social Research	Working on the process evaluation

TIMELINE

Date	Activity	Organisation
September 2013 - May 2014	Recruit schools	Achieve Together
September 2014	Year one of the Achieve Together pilot (for Cohort 2) begins. Year two of the Achieve Together pilot (for Cohort 1) begins.	Achieve Together
November 2014	Telephone interview with the Achieve Together project manager and school leads in five Cohort 2 case study schools.	NatCen
May –July 2014	Case study visits to five Cohort 2 schools	NatCen
September 2015	Year two of the Achieve Together pilot (for Cohort 2) begins. Year three of the Achieve Together pilot (for Cohort 1) begins.	Achieve Together
October 2015	Apply for access to NPD for KS4 results for June 2013, June 2014, June 2015 (year one of the pilot for Cohort 2), absences and prior KS2 attainment.	IFS
November 2015	Analysis KS4 attainment and absences after the first year of the pilot for Cohort 2 schools, and comparison of impact estimates for Cohort 1 schools to inform the likely bias for Cohort 2 schools.	IFS
April 2016	Presentation of interim findings.	IFS
May-July 2016	Case study visits to five Cohort 2 schools	NatCen
October 2016	Collect KS4 results for June 2016 (year two of the pilot for Cohort 2).	IFS
November 2016	Analysis KS4 attainment and absences after the second year of the pilot for Cohort 2 schools.	IFS
30th January 2017	Draft final report	IFS / NatCen
May 2017	Final report	IFS / NatCen

Risks

IFS is registered under the Data Protection Act 1998 (registration number Z5758698) and complies with all its obligations. IFS also ensures that its staff and anyone else involved in its work abide by its Data Security Policy which details the measures that are in place to protect data and to ensure compliance with any legal requirements.

Risk	Likelihood	Impact	Mitigation and contingency plan
Sample sizes too small to detect significant impacts	High	High	Given the current sample sizes, our power calculations show we would be able to detect small effect sizes if there is no intra class correlation (ICC). However, if the ICC were 0.1 or higher (very plausible), we would only be able to detect effect sizes of 0.2 or higher. This effect size is larger than most estimates of the standard deviation of teacher effectiveness. This is a high risk and can only be mitigated by higher samples sizes. If results are insignificant, we will report the estimated effect, p-value and confidence interval.
Unavailability of staff during project	Low	Low	IFS has a pool of researchers on hand with expertise in programme evaluation, education policy and qualitative research methods. Systems and procedures will be adequately documented to ensure handover can occur smoothly if necessary.
Delay in access to data	Low	Medium	We have planned for GCSE results to be made available in the Spring of each year (as is currently the case). This could change over time as DfE review their arrangements.
Loss of or damage to data	Low	High	IFS has high levels of IT security in place. All members of the evaluation team have extensive experience of working with data, and are very aware of the importance of keeping data safe and of using the necessary security procedures. Back-ups are located off-site and can be retrieved within one working day. EEF and the project team will be immediately notified if data is accidentally damaged or stolen; contingency plans specified in our security policies (available on request) can then be put into place.
Matched comparable group is not credible	Medium	High	We will compare the impact estimates from the first year of Achieve Together in the first cohort when using a matched design and when using the second cohort as a control group. The comparability of these impact estimates will inform whether the matched design is credible. We are reasonably sure that the matched design will be sufficiently unbiased to use this methodology to estimate the impact of the second cohort involved with Achieve Together.
Lack of engagement from case study schools	Low	High	The process evaluation will depend on schools engaging with the research process and feeding their views and experiences back at various points throughout the two years. To maximise school engagement, the research team will minimise burden by conducting some interviews by phone and offering long fieldwork periods and flexible appointment times for case study visits.