Education Endowment Foundation

Community Apprentice

Evaluation report

October 2024

Dr Patrick Taylor, Kimberly Bohling, and Lilli Wagstaff

THE BEHAVIOURAL INSIGHTS TEAM

The Education Endowment Foundation (EEF) is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries, and colleges to improve teaching and learning for 2 to 19-year-olds through better use of evidence.

We do this by:

- **summarising evidence**: Reviewing the best available evidence on teaching and learning and presenting in an accessible way;
- **finding new evidence**: Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socioeconomically disadvantaged backgrounds; and
- **putting evidence to use**: Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

This project was co-funded by The Careers & Enterprise Company and the Bank of America.

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

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

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

This project was evaluated by an evaluation team at the Behavioural Insights Team.

The project was led by Dr Patrick Taylor and Kimberly Bohling. The impact evaluation was designed by Hazel Wright, Dr Patrick Taylor, Kimberly Bohling, Pantelis Solomon, and Michael Sanders. Analysis was conducted by Tom Middleton, Martina Maglicic, and Lilli Wagstaff. Survey data collection was managed by Louise Jones and Stephanie Petrakis and carried out by a team of field research assistants employed by BIT. The IPE was designed by Jessica Heal. IPE field data collection was conducted by Doireann O'Brien, Patrick Taylor, Kimberly Bohling, and Ingrid Broch-Due.

Contact details:

The Behavioural Insights Team

58 Victoria Embankment,

London,

EC4Y 0DS

Email: info@bi.team

Acknowledgements

Envision: Jennie Butterworth, Hannah Matthews, Dan Heffernan, Ben Harding, and Jenny Welsh.

Education Endowment Foundation: Guillermo Rodriguez, Triin Edovald, Rachael Morris, Eleanore Widger, Belén Parada Zuleta.

This project was co-funded by The Careers & Enterprise Company and the Bank of America.

Executive summary

The project

Community Apprentice is a youth social action programme in which teams of Year 10 pupils (14 to 15-year-olds) are supported to design and deliver a fundraising campaign for a charity of their choice. The programme aims to develop pupils' self-efficacy, social confidence, and teamwork skills, with the long-term aim of increasing their participation in social action. The programme (which has been redesigned and further codified since being delivered as part of the Education Endowment Foundation [EEF] trial) is delivered by Envision, with support from trained local business volunteers. The programme takes place over 12 weeks and includes: a full-day cross-school launch event; 11 weekly sessions in school, facilitated by an Envision coach; four business mentoring sessions hosted by a business mentor volunteer; at least one awareness raising event hosted by the pupils in school; and a final half-day 'Apprentice-style' cross-school event, where teams pitch their fundraising projects to local business leaders.

Community Apprentice was evaluated using a two-arm randomised controlled trial, with randomisation at pupil level. The two arms were treatment and control, with pupils randomly assigned to the treatment group offered the Community Apprentice programme. Around 761 pupils from 30 schools were randomly allocated to the treatment or control group. Around 47% of randomised pupils were eligible for free school meals (FSM), exceeding the 30% target set in the evaluation design. In line with the EEF mission, the primary outcome measure was academic attainment, as measured by the General Certificate of Secondary Education (GCSE) Attainment 8 scores. While the programme does not aim to improve academic attainment, a logic model was developed that described how the programme might achieve this.

An implementation and process evaluation (IPE) was carried out alongside the impact evaluation, which sought to understand intervention delivery and participant engagement. The IPE involved interviews, observations, surveys, and administrative data analysis. Interviews and observations were conducted in three case study schools. One programme activity was observed in each of these schools. In all case study schools, interviews were carried out with pupils and a selection of staff and volunteers involved in the delivery or management of the intervention. The trial started in October 2017 and pupils sat their GCSE exams in July 2019. Analysis and reporting were completed in November 2023 due to delays in accessing the National Pupil Database.

Envision, founded in 2000, aims to develop young people's confidence and soft skills by empowering them to tackle issues within their community. The trial was co-funded by The Careers & Enterprise Company and the Bank of America.

Table 1: Key conclusions

Key Conclusions	
1.	Pupils in the Community Apprentice intervention group made no additional progress in their GCSE Attainment 8 score compared to pupils in the control group. This is our best estimate of impact, which has a moderately high security rating. However, there is a high level of uncertainty around the result: the possible impact of the programme could range from a negative effect of two months' less progress to a positive effect of two months' additional progress.
2.	Pupils in the intervention group showed improvements in employability skills and attitudes, particularly self-efficacy and social confidence, compared to pupils in the control group. While there is some uncertainty around effect sizes for these outcomes, findings suggest that the programme is likely to have had a positive impact on self-efficacy and social confidence, in line with the programme's aims.
3.	Subgroup analysis on whether the programme was more or less effective for FSM-eligible pupils is inconclusive due to the small sample size.
4.	The programme was mostly delivered as intended, suggesting that the estimated effects are likely to be due to the programme design. However, the IPE found a high level of variation in the quality of delivery by coaches and mentors. Envision coaches were found to have an important impact on quality, especially their ability to strike a balance between directive and non-directive session facilitation. More consistency in the quality of coaching would likely lead to greater positive effects.
5.	The logic model hypothesised that the programme would have a positive effect on pupils' attitudes towards their education. No evidence was found to support this, which might explain the lack of effect on academic attainment.

EEF security rating

These findings have a moderately high security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in 30 schools. The trial was a well-designed two-armed randomised controlled trial, with randomisation at the pupil level to maximise power and a minimum detectable effect size (MDES) of 0.19, which complies with the EEF threshold to receive the maximum-security rating. Missing data for the primary outcome (GCSE Attainment 8 score) was relatively low (9%). The security rating was adjusted down one padlock due to some

threats to internal validity that could have diluted the effects of the intervention and make it harder to accurately estimate the size of the impact on the pupils in the trial. Notably, the intervention was not designed to impact attainment, so the trial was likely underpowered to find an effect on the primary outcome measure. Moreover, GCSE Attainment 8 was collected over a year after the end of the intervention, which might also explain the absence of a treatment effect.

Additional findings

The impact analysis suggests that the programme had a substantial positive effect on the secondary outcomes, showing improvements in pupil self-efficacy and social confidence. The IPE provides strong evidence that programme activities were delivered as intended, although variation was found in the quality of delivery. Findings from interviews and observations supported the mediating mechanisms hypothesised in the logic model, showing that pupil ownership, the completion of mini challenges, achieving goals, winning the competition, and forming positive relationships with teammates supported pupil engagement and contributed to the secondary outcomes.

The evidence from this study suggests that the hypothesised moderating factors were less well identified in the logic model. Four moderating factors emerged as important, providing insight into how the intervention works: the quality of the coach; baseline pupil characteristics; team dynamics; and school characteristics. The IPE also identified some potential unintended mechanisms and outcomes. As well as the mechanisms identified in the logic model, the effects of the intervention may have been driven by pupils experiencing controlled discomfort, vicariously experiencing changes in confidence in peers, taking part in activity perceived as difficult, and one that had 'real-world' applications. In addition to the hypothesised outcomes, in interviews, pupils reported that they felt an increase in their action-orientation, resilience, organisation skills, and participation in future social action. Both the hypothesised and newly identified mechanisms seem dependent on the group and project-based nature of the programme. No significant difference was found between the intervention and control groups in terms of their participation in other extracurricular activities, which supports the conclusion that the effects estimated in the impact analysis are due to the intervention.

These results build on the existing literature in two main ways. First, they add robust evidence to the wider literature and for the EEF Teaching and Learning Toolkit—on the effects of youth social action on academic performance. Second, they add to the growing body of literature that identifies a positive causal effect of youth social action programmes on non-cognitive skills and attitudes (Cameron *et al.,* 2017; Fitzpatrick *et al.,* 2021; Kirkman *et al.,* 2016). This study suggests that the largest effects are on young people's social confidence and self-efficacy; a very similar finding to the latest academic work on the topic that combined quasi-experimental impact findings with in-depth qualitative research to build a theory of how youth social action works (Taylor, 2021).

Cost

The total cost per school per year of delivering the intervention is £5,970.76 (equivalent to £459.29 per pupil per year). The intended business model for the intervention in the future includes a fee to schools of £500 per year, with the remaining costs incurred by Envision and paid for by trusts, foundations, and corporate partners. Outside of the evaluation, it is therefore, estimated that the cost will be £604.10 per year (or £46.47 per pupil per year).

Impact

Table 2: Summary of impact on primary outcome(s)

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P-value	EEF cost rating
GCSE Attainment 8	0.0067 (-0.096, 0.11)	0	🔒🔒🔒🔒🔓	689	0.90	£ £ £ £ £

Introduction

Background

Existing evidence

The intervention being evaluated is a programme of 'youth social action'. Youth social action is defined as activities in which young people are supported to take 'practical action in the service of others' (Pye and Michelmore, 2017). Youth social action ranges from informal volunteering to structured programmes of learning. High-quality social action is believed to generate a dual benefit: participating individuals can develop skills, knowledge, attitudes, and relationships; and the wider community can benefit from the actions that they take. Evidence for this dual benefit is steadily accumulating, with the strongest evidence in support of a benefit to participating young people (Birdwell *et al.,* 2015).

Programmes of youth social action are increasingly seen as a form of careers education that may increase young people's future employability (Birdwell *et al.,* 2015). A longitudinal study conducted by the United States Corporation for National and Community Service reported volunteers in programmes like AmeriCorps as being 27% more likely to find employment than those who had not volunteered (Spera *et al.,* 2013). The impetus for youth social action programmes to try to increase and measure employability came partly in response to feedback from employers and other organisations that schools do not sufficiently cultivate the skills necessary to succeed in today's competitive labour market (Birdwell *et al.,* 2015). Landmark research conducted by Nobel Laureate James Heckman and Tim Kautz indicates that when it comes to labour market success, attributes such as conscientiousness, which is linked to self-control and perseverance, and agreeableness, which is linked to characteristics like empathy, modesty, and trust are as important as educational attainment (Heckman and Kautz, 2013).

There is growing empirical evidence to suggest that participation in youth social action programmes can help young people develop non-cognitive skills and positive character traits. Participation in youth social action may, for instance, spur employability by increasing self-efficacy in a range of domains. Evaluations of social development and community action initiatives such as the National Citizens Service (NCS) indicate positive impacts of involvement on participants' self-belief that they have the skills and experience to attain employment in the future, as well as their social self-efficacy ('an individual's belief in her social interaction and communication skills, and her belief that applying these skills can achieve positive outcomes' (Taylor, 2021, p. 47); a construct very close to one of the secondary outcomes in this present study). Increases in participants' social self-efficacy, their confidence in their ability to get things done on time and to try new activities were also reported (Cameron *et al.,* 2017; Fitzpatrick *et al.,* 2021; Taylor, 2021). Research conducted by the Behavioural Insights Team (BIT), including the evaluation of an Envision programme very similar to the Community Apprentice intervention, found significant increases in a range of outcomes including attitudes to education and grit (Kirkman *et al.,* 2016). These non-cognitive skills are thought to underpin achievement in both school and work environments (Gutman and Schoon, 2013).

There is some limited evidence to support the idea that participation in youth social action programmes may also influence academic outcomes. One mechanism through which youth social action could help to improve attainment is by enhancing young people's attitudes towards school, and by motivating students to apply themselves to their studies to achieve better grades (Birdwell *et al.,* 2015). Findings from the Effective Provision of Pre-school Education (EPPE) Project demonstrate that self-reported enjoyment of school at age 11 is associated with higher attainment in Maths at Key Stage 3 assessments (Sylva *et al.,* 2012).

Another way in which youth social action may help to improve educational outcomes is by enhancing young people's self-confidence and well-being. Evidence from the 2013 Demos report Service Nation demonstrated that 80% of social action programme participants reported increases in their sense of self-confidence (Birdwell *et al.,* 2015). Gutman and Schoon's (2013) review on the impact of non-cognitive skills on youth outcomes indicates that adaptive coping emotions such as confidence have been positively associated with achievement in academic settings. A comprehensive report by the Department for Education (DfE) analysed the relationship between different dimensions of well-being and educational outcomes (using national exam scores) (Gutman and Vorhaus, 2012). Higher levels of emotional, behavioural, social, and school well-being were found to be correlated with higher levels of academic achievement and school engagement at both ages 7 to 13 and ages 11 to 16. This analysis also indicated that over time, these dimensions of well-being become more influential in explaining school engagement than demographic variables. In terms of older

cohorts, an evaluation of the NCS reported that programme involvement increased self-reported intention to study for a degree or other higher education qualification among those aged 15–17 (Cameron *et al.,* 2017).

The only UK-based evaluation of youth social action's direct effect on academic attainment prior to this present study was a large-scale randomised controlled trial, funded by the Education Endowment Foundation (EEF) (Gorard *et al.,* 2016). This trial investigated whether participation in school-based uniformed youth organisations such as the Scouts or the Fire Cadets would increase academic performance at Key Stage 3. Overall, the results of the trial were inconclusive. The findings indicated no significant effect on attainment. However, data access issues meant that internal school assessment records at the end of Year 9 were the only available proxy for Key Stage 3 scores. These internal school records only weakly correlate with standardised test scores, undermining the reliability and validity of this finding.

The evidence of impact on attainment of social and emotional learning (SEL) programmes more broadly (as opposed to youth social action specifically) is positive. A meta-analysis of school-based SEL programmes found positive average effects on reading (an effect size of +0.25 standard deviations [SDs]), mathematics (effect size of +0.26), and science (effect size of +0.19) (Corcoran *et al.,* 2018, p. 61). The three EEF-funded randomised controlled trials of SEL programmes (which are not included in the Corcoran *et al.* 2018 meta-analysis) found no impact on academic attainment (Humphrey *et al.,* 2015; Wiglesworth *et al.,* 2018; Humphrey *et al.,* 2018), but the Corcoran *et al.* (2018) meta-analysis is the most robust overall assessment of impact of SEL identified in this review. It covered 611 studies from 1970 to 2016, with 40 meeting sound inclusion criteria that allowed only for randomised controlled trials and robust quasi-experimental designs (Corcoran *et al.,* 2018, p. 59). The programmes included in the review had to give 'children and adults [the opportunity to] acquire and effectively apply the knowledge, attitudes, and skills necessary to understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions' (Collaborative for Academic, Social, and Emotional Learning, 2015, p. 5). This covered a range of programmes that share some characteristics with youth social action—often with group-based and project-based activities, and with similar target outcomes—but also differ in many aspects (e.g. most do not focus on social action), and cover a diverse set of activities (usually focused more explicitly on strategies for identifying and managing emotions). An exposition of these differences is beyond the scope of this brief review but is described in detail in the meta-analysis. Harvard University's Taxonomy Project also provides helpful resources to help researchers and practitioners to unpick the differences between different types of SEL-related programmes and outcomes.[1]

Policy and practice relevance

A substantial amount of public resources have been invested in youth social action in the 2010s. In England, successive rounds of government funding (the 'Youth Social Action Funds') supported a range of youth social action organisations, including Envision, to grow. The single largest recipient of government support has been NCS. In 2015, the UK government committed over £1bn to NCS over five years, with an estimated cost per participant of £1,863 (National Audit Office, 2017). Similar policies have been proliferating across Europe, through programmes led by the European Union and by individual states (Zimenkova, 2013), and the idea is well-established in the United States (under the banner of 'AmeriCorps') and elsewhere. This has been a controversial policy in England; particularly as it has coincided with a decline in government support for formal citizenship education in schools (Kerr, 2014).

Rationale for this evaluation and its design

Given the current paucity of evidence, this evaluation of the Community Apprentice programme sought to determine whether improvements in academic outcomes can be achieved, alongside three critical skills for employment (general self-efficacy, social confidence, and teamwork) through participation in a programme of youth social action. A randomised controlled trial design was used to estimate the effects of participating in Envision's Community Apprentice programme (treatment arm) versus 'business as usual' (control arm) for pupils in participating schools. Randomisation was carried out at the pupil level to maximise the power of a fixed sample that was determined by the developer's delivery capacity. This randomised controlled trial was complemented by an implementation and process evaluation (IPE) that sought to understand intervention delivery and participant engagement, to support interpretation of the findings from the impact evaluation.

[1] https://easel.gse.harvard.edu/taxonomy-project.

Intervention

Why

This intervention, called the Community Apprentice, aims to develop participants' sense of general self-efficacy, social confidence, and teamwork skills. It does this by supporting teams of Year 10 pupils to design and deliver a fundraising campaign for a charity chosen by them, in the workplace of a local business. The long-term goal of the programme is to increase participants' participation in social action. It is also hoped that this experience will result in a more positive attitude towards education, leading to an improvement in academic attainment. Academic attainment is the primary outcome of interest in this evaluation, but the intervention was not designed to improve attainment. The decision to take attainment as the primary outcome was based on the EEF mission to break the link between family income and educational achievement, and a desire to understand whether programmes like the Community Apprentice can help to achieve this. The design of the intervention was set before the research team's engagement with the EEF and was not modified during the evaluation's set-up phase. As part of their application for funding to the EEF, Envision had to provide a plausible rationale for how the intervention could increase academic attainment. This rationale was further developed into a logic model with support from BIT during the set-up phase of the evaluation. The initial logic model therefore, represents a description of how the intervention could achieve the EEF core aim, as opposed to Envision's core aims for the programme.

What

The programme combines project work in small teams (13 pupils) with mentoring by local business leaders and facilitation provided by an Envision coach. Teams participate in a combination of in-school sessions, business-hosted sessions, and cross-school events. The results of pupils' fundraising campaigns are judged as part of a cross-school competition, which also considers their personal development during the process (assessed in terms of the number of 'skills badges' they earn).

Who

Delivery of the intervention in each school involves four key roles as follows:

- *Envision coaches*: Each team of pupils is assigned an Envision coach. This is a member of staff from Envision with lead responsibility for supporting the team. They facilitate the weekly meetings and business mentoring sessions and work with the lead teacher to support project implementation.
- *Business mentors:* Envision recruits staff from local companies to volunteer to support each school team. These volunteers offer advice through the business mentoring workshops, support the pupils' fundraising by hosting a fundraising activity at their offices, and support the team by attending the 'Boardroom final'.
- *Lead teacher:* This is the member of staff at each host school that coordinates the intervention and accompanies them to the external events. They may also attend the weekly meetings facilitated by the Envision coach, though this is not compulsory. Occasionally this role is shared by multiple members of staff and it is sometimes fulfilled by non-teachers (e.g. teaching assistants or careers coordinators). They are selected by the school's senior management during the school recruitment process.
- *Senior manager:* This is the member of senior staff at each school with overall responsibility for the implementation of the intervention in their school.

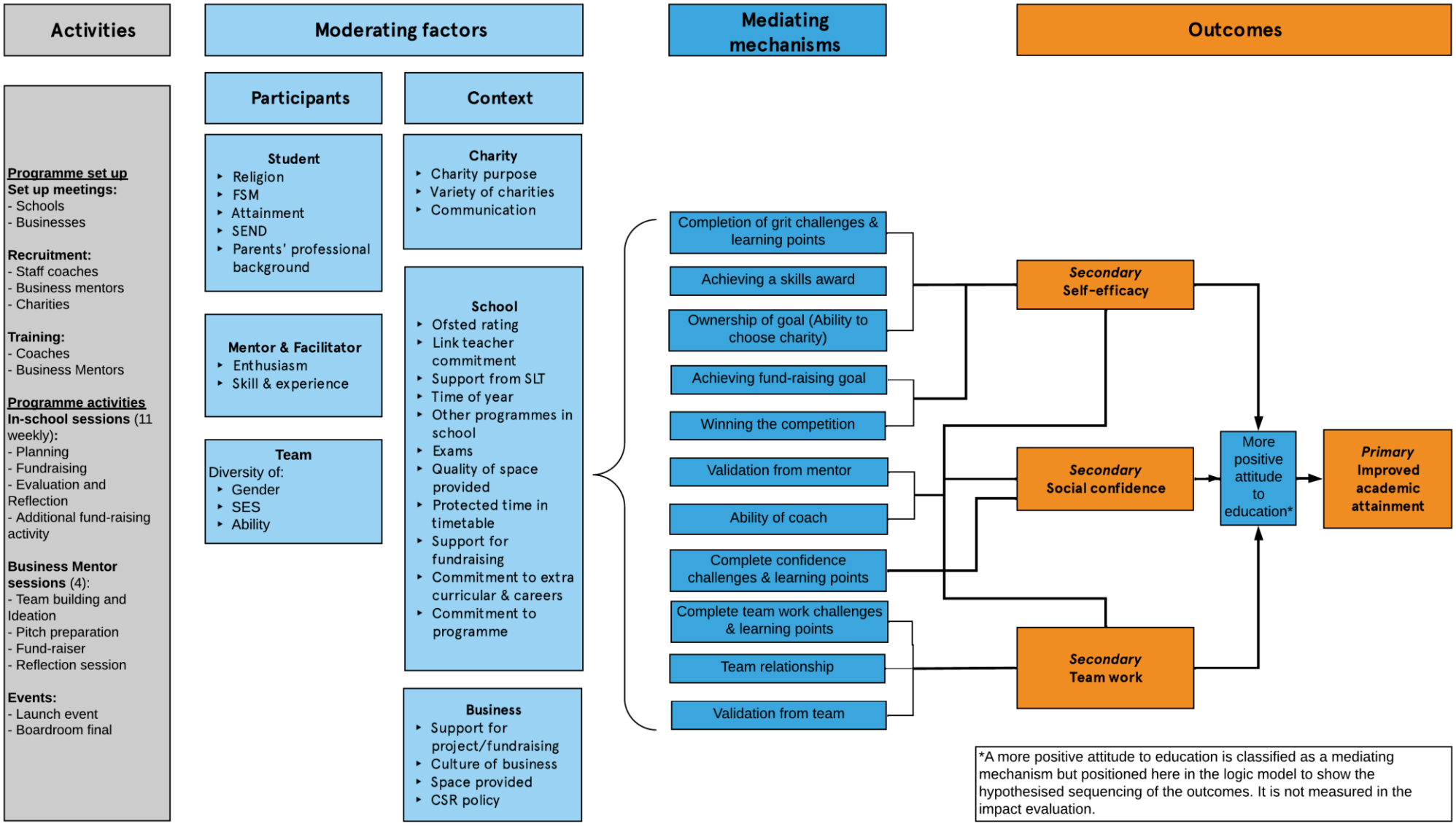
How, where, when, and how much

The intervention is delivered by Envision, an organisation founded in 2000. Envision aims to develop young people's confidence and non-cognitive skills by empowering them to tackle issues within their community. The programme takes place over a 12-week period and includes the following five elements:

- *Cross-school launch event:* One six-hour cross-school event at a local venue that exposes pupils to a variety of local charities who pitch for their support. At this event, pupils choose which charity they will support with the fundraising project.
- *Weekly meetings:* Eleven one-hour sessions in school, in which the pupils are supported by Envision coaches to develop plans for their fundraising event and develop and rehearse presentations. While doing this they earn points for their team by undertaking a range of personal skills challenges.
- *Business mentoring:* Four two-hour business mentoring sessions, in which teams attend workshops with trained volunteers from local companies to further develop their plans. These sessions are facilitated by Envision to build social confidence and skills.
- *Project implementation:* Pupils implement an awareness raising event in the school and a fundraiser in their partner business. These are short one-off activities, which take place on one day and generally last an hour or two.
- *Cross-school 'Boardroom final':* A final three-hour 'Apprentice-style' cross-school event at a local venue in which teams 'pitch' against rival schools to a panel of local business leaders, presenting the results of their fundraising campaigns and describing their own personal development. The judges select the Community Apprentice team of the year at this event.

The two cross-school events take place during the school day. School-based weekly meetings take place either during lunch break, directly after school, or during curriculum time; a choice that is made by the school. Business mentoring sessions take place at the end of the school day, or directly after school. **Figure 1**, below, shows the logic model that was developed with Envision prior to the evaluation. Further details of the intervention can be found in **Appendix D**.

Figure 1: Logic model



Evaluation objectives

Impact evaluation

The impact evaluation had one primary research question and seven secondary questions.[2]

Primary research question:

1. What is the effect of being assigned to the programme on overall GCSE (General Certificate of Secondary Education) performance, as per the Attainment 8 score at the end of Year 11?[3]

Secondary research questions:

1. What is the effect of being assigned to the programme on English attainment, as assessed by GCSE assessments taking place at the end of Year 11 (English scores averaged across English Language and English Literature)?
2. What is the effect of being assigned to the programme on Maths attainment, as assessed by GCSE assessments taking place at the end of Year 11?
3. What is the effect of being assigned to the programme on teamworking skills, as indicated by the 'Teamwork Scale for Youth (TSY)'?
4. What is the effect of being assigned to the programme on social confidence, as indicated by the 'Self-Perceived Communication Competence Scale (SPCCS)'?
5. What is the effect of being assigned to the programme on self-efficacy, as indicated by the 'New General Self-Efficacy (NGSE) Scale'?
6. What is the effect of being assigned to the programme on GCSE performance for pupils eligible for free school meals (FSM; variable Ever 6 FSM), as measured by Attainment 8 and GCSE point scores in both Maths and English? (English scores averaged across English Language and English Literature).
7. What is the effect of being assigned to the programme on pupils' progression into further social action activity, as measured through participation in the NCS?[4]

The evaluation protocol and Statistical Analysis Plan can be found on the EEF website.

Implementation and process evaluation (IPE)

The IPE addressed the following questions:

1. Programme delivery:
    a. To what extent is the programme delivered as intended (fidelity and dosage)?
        i. To what extent do the young people get the exposure to the intervention as planned?
        ii. To what extent do coaches and business mentors adhere to the intended treatment model?
        iii. How much—and what kind of—variation is there across schools? (adaptation)
    b. How well is the programme delivered (quality)?
        i. Do the young people receive positive reinforcement and validation from the business mentors? What impact does this have on the three secondary non-cognitive outcomes?

[2] These questions have been slightly refined from the protocol to make it clear that the primary analysis is on an intention-to-treat (ITT) basis with two-sided hypothesis tests. As a result, the phrasing refers to 'assignment to the programme' (instead of 'participation in the programme') and asks 'what is the effect?' (instead of asking whether there is a positive effect).
[3] Attainment 8 measures pupils' average grade across eight subjects at the end of year 11 (summer 2019).
[4] This question has not been answered as the NCS was unable to share the necessary data.

   ii. Do the coaches' abilities impact delivery?
c. How distinct is the programme from existing practice (programme differentiation)?
   i. Do the schools offer any other programmes of a similar nature (e.g. fundraising and careers engagement)?
   ii. Are the pupils in the control group engaged in similar activities (monitoring of the control group)?

2. Programme engagement:

   a. To what extent is the programme delivered to its intended recipients (reach)?
      i. How many of the intervention group students are eligible for FSM (as defined by the Ever 6 FSM variable) (in line with required 30%)?
      ii. What motivates students to participate or not?
      iii. What are the barriers/facilitators to student participation?
   b. To what extent are recipients engaged during the delivery of the programme (responsiveness)?
      i. To what extent do the participants meaningfully engage with the programme? What impact does completing 'challenges' or 'learning points' have? Do they feel that they have ownership of their goal, and what impact does this have?
      ii. How does the relationship within the young people's team contribute to engagement?
      iii. What impact does winning the 'Boardroom final' have?
      iv. To what extent does the school engage in the programme?

3. Programme theory:[5]

   a. What is the perceived impact of the programme?
   b. What are the key mediating mechanisms?
   c. What are the key moderating factors?

Ethics and trial registration

The project was subjected to and passed BIT's internal ethics review. The review concluded that the consent procedures were appropriate for each type of data gathered and that appropriate data protection protocols were in place. There were no concerns about coercion or potential harm to pupils.

Schools that agreed to participate in the trial signed a Memorandum of Understanding (MOU), which clearly outlined the aims of the intervention and evaluation, as well as the responsibilities of Envision and the school (see **Appendix E**).

The trial registry number is ISRCTN27341369.

Data protection

This evaluation commenced in Autumn Term 2017, so all initial data collection and storage procedures were in compliance with the Data Protection Act of 1998. Under that legislation, using consent as the lawful basis for processing required only opt-out consent. Schools were each asked to identify 26 students who were interested in participating in the programme and provide them with an information sheet about the aims of the project and evaluation, as well as how their data will be used. The information sheet provided an opportunity for parents to opt their child out of the programme and study.

The Data Protection Act of 2018 (the General Data Protection Regulation or GDPR) came into effect when the programme was nearly complete and secondary outcome data was being collected (May 2018). The GDPR sets a higher standard for processing under the lawful basis of consent and requires explicit, opt-in consent. Accordingly, the consents we obtained previously would not, if we implemented under the GDPR, be sufficient to satisfy the new requirements. However, the GDPR does not require that consent obtained under the Data Protection Act of 1998 be 'refreshed' in all cases. The Information Commissioner's Office guidance on this topic indicates that in cases where consents obtained under the Data Protection Act of 1998 are not sufficient to satisfy the regime under the GDPR, processing could take

[5] These questions were not included in the original protocol. However, given the lack of strong evidence to support a lot of the assumptions in the intervention logic model, it was decided—prior to IPE data collection—to add some exploration of the theory. This has allowed for a refinement of the logical model that is presented in the findings.

place under a different lawful basis.[6] If we had launched the trial under the new GDPR, given the practical issues with obtaining opt-in consent we would have instead relied on the lawful basis of legitimate interests. In this case, we would have had schools send home information sheets providing the same information that was provided in the opt-out consent form with the same opportunity to withdraw their student's data at any time by returning a form or contacting the evaluation team.

In the present case, given the passage of time and the changeover in school cohorts it would be very difficult to obtain opt-in consent. Hence, we cannot rely on that basis under the GDPR and instead look to the legitimate interest basis. As we have already given students and parents an opportunity to object to inclusion in the study and made clear that the subjects of the study are able to object to their inclusion at any stage, we have in all practical terms already completed what we would otherwise have been required to do under the GDPR in order to satisfy the requirements of the legitimate interest basis for data processing.

After balancing the interests of the subjects of the study, we relied on the lawful basis of legitimate interests under the GDPR and processed the data relating to the study as originally envisaged without the need to obtain any further consent from or provide further information to parents/students. BIT is the data controller of all data related to the evaluation up to the point that the data is transferred to the archive, at which point the EEF becomes the controller.

However, as an additional precaution to ensure we were able to collect the necessary data from the National Pupil Database (NPD), we distributed an updated Data Processing Notice (DPN) (see **Appendix F**), which informs parents/guardians of their rights in regard to data processing for this trial. This DPN was distributed as part of the process in which we distributed new information sheets (see **Appendix G**) informing parents and students about their opportunity to withdraw from data processing in relation to the longitudinal outcomes, as these outcome measures were not specified in the original opt-out consent form that parents received.

Project team

The intervention implementation was led by the following staff at Envision: Jennie Butterworth, Chief Executive Officer; Hannah Matthews, Quality and Impact Manager; Dan Heffernan, Birmingham Manager; Ben Harding, Bristol Manager; and Jenny Welsh, London Manager.

The evaluation was delivered by the following staff at BIT. The project was led by Dr Patrick Taylor and Kimberly Bohling. The impact evaluation was designed by Hazel Wright, Dr Patrick Taylor, Kimberly Bohling, Pantelis Solomon, and Michael Sanders. Analysis was conducted by Tom Middleton, Martina Maglicic, and Lilli Wagstaff. Survey data collection was managed by Louise Jones and Stephanie Petrakis and carried out by a team of field research assistants employed by BIT. The IPE was designed by Jessica Heal. IPE field data collection was conducted by Doireann O'Brien, Patrick Taylor, Kimberly Bohling, and Ingrid Broch-Due.

[6] https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/consent/whats-new.

Methods

Trial design

Table 3: Trial design

Trial design, including number of arms		Randomised controlled trial with two arms
Unit of randomisation		Individual
Stratification variable(s) (if applicable)		School and FSM status
Primary outcome	Variable	Academic attainment
	Measure (instrument, scale, source)	GCSE Attainment 8 score, 1–9, NPD
Secondary outcome(s)	Variable(s)	Maths attainment
	Measure(s) (instrument, scale, source)	GCSE Maths point score, 1–9, NPD
Secondary outcome(s)	Variable(s)	English attainment
	Measure(s) (instrument, scale, source)	GCSE English point score, 1–9, NPD
Secondary outcome(s)	Variable(s)	Self-efficacy
	Measure (instrument, scale, source)	New General Self-Efficacy Scale, 1–5, bespoke survey
Secondary outcome(s)	Variable(s)	Teamwork
	Measure (instrument, scale, source)	Teamwork Scale for Youth, 1–5, bespoke survey
Secondary outcome(s)	Variable(s)	Social confidence
	Measure (instrument, scale, source)	Self-perceived Communication Competence Scale, 1–5, bespoke survey
Baseline for primary outcome	Variable	Academic attainment
	Measures (instrument, scale, source)	Key Stage 2 raw Reading scores, 0–50, NPD Key Stage 2 raw Maths scores, 0–110, NPD
Baseline for secondary outcome(s)	Variable	Reading attainment
	Measure (instrument, scale, source)	Key Stage 2 raw Reading scores, 0–50, NPD
Baseline for secondary outcome(s)	Variable	Maths attainment
	Measure (instrument, scale, source)	Key Stage 2 raw Maths scores, 0–110, NPD

This was an efficacy trial conducted as a two-arm individually randomised controlled trial, with randomisation at the pupil level. The two groups are treatment and control. Pupils were randomly assigned to either the Community Apprentice programme (treatment group) or ‘business as usual’ (control group). An indication of what constituted ‘business as usual’ is given in the ‘Programme differentiation’ subsection in the ‘IPE results’ section below.

Our primary outcome is academic attainment, measured using the GCSE Attainment 8 score. As secondary outcomes, Maths and English attainment were analysed separately, measured with the GCSE Maths point score and GCSE English point score, respectively. In addition to attainment, there are three other secondary outcomes: self-efficacy; teamwork; and social confidence.

No changes were made to the trial design specified in the trial protocol and Statistical Analysis Plan.

Participant selection

The trial was conducted with Year 10 pupils in state secondary schools in Birmingham, Bristol, and London (the three operating locations of Envision). Schools that met the following four criteria were eligible to participate in the trial.

- *Location:* Schools had to be based in Birmingham, Bristol, or London.
- *Feasibility of delivering the intervention:* Teaching and management staff were expected to have the capacity to support the administration of the project and the flexibility in their timetable to implement the Community Apprentice programme.
- *Pupil disadvantage:* Schools had to agree that it was feasible to nominate 30% of pupils to the trial who were eligible for FSM (Ever 6 FSM).[7] Data containing FSM status for each pupil was collected prior to randomisation (details below).
- *Consent procedures:* Schools had to agree with the opt-out consent procedure that the research team planned to use with pupils and parents.

Envision recruited 30 schools that met these four eligibility criteria.

Each school was then responsible for selecting 26 pupils (to allow for 13 for in each trial arm, see 'Randomisation' section for more details on how allocation took place) who met the following two eligibility criteria.

1. Year group: Pupils had to be in Year 10.
2. Pupil disadvantage: Schools had to aim to select pupils such that 30% of the sample were eligible for FSM (Ever 6 FSM), but the school was not excluded from the study if it fell short of this percentage.

Envision did not believe that it would be possible to place any further constraints on the pupil selection process due to the timings of the trial set-up and the newness of the intervention.

This gave a total sample size of 761 pupils (390 in the intervention group and 371 in control). Pupils in each school who were randomly allocated to the control group continued with 'business as usual' activity for the duration of the trial.

Schools wishing to participate in the trial were asked to sign an MOU (see **Appendix E**) before enrolment, agreeing to the required activities for both the intervention and evaluation. This included the requirement for each school to provide a named 'school project coordinator' to ensure MOU commitments were met.

As presented in the MOU, schools were responsible for seeking opt-out consent from pupils and from the parents (or legal guardians) of all eligible pupils. However, four schools switched to an opt-in process on their own initiative. With either process, consenting participants and their parents (or legal guardians) agreed to pupil participation in the study, randomisation into treatment or control arms, and the sharing of the Unique Pupil Number (UPN) and other personal pupil data with BIT. After removing those who did not consent (ten students), schools shared pupil-level data with BIT for the purpose of randomisation. Pupils were informed of their trial arm allocation by December 2017.

[7] This was a funding requirement imposed by the EEF in line with its mission.

Outcome measures

Baseline measures

The research team used Key Stage 2 Maths and Reading scores obtained from the NPD as control variables in our analyses. These were chosen based on the high correlation between Key Stage 2 scores and GCSE results (Key Stage 4 scores).[8] These were thought to be the most appropriate measures given Key Stage 3 results were not available for our sample.

Primary outcome

The primary outcome is overall GCSE performance, as measured by the Attainment 8 score obtained from the NPD. This was assessed at the end of Year 11 (summer 2019, one year after completion of the intervention). It is the sum of the points for a pupil's eight best GCSE scores from a set of DfE-approved GCSE subjects, with a double-weighting applied to English and Maths. This combined measure was chosen because the intervention developers believed that any effect that the intervention had on academic attainment would not be domain-specific. As shown in the logic model, they hypothesised that the programme might lead to a generally more positive attitude to education, which could in turn lead to a general improvement in academic attainment. Attainment 8 is the main combined attainment metric used by the DfE to assess school performance, so has particular importance in education policy in England.

Secondary outcomes

The evaluation also estimates the effects on five secondary outcomes. Two secondary outcomes are Maths and English attainment separately:

- English attainment, as assessed by GCSE assessments that took place at the end of students' Year 11 (summer 2019). Both English GCSEs point scores were used to create an average final English score.
- Maths attainment, as assessed using GCSE Maths point score from the Year 11 (summer 2019) GCSE assessment.

Three secondary outcomes are non-academic, employability skills, and attitudes:

- Self-efficacy, as assessed by the "New General Self-Efficacy scale" (NGSE Scale (Chen *et al.,* 2001). This is a seven-item self-report survey that contains items relevant to pupils' sense of self-efficacy. Participants were asked to respond on a scale of 1 (strongly disagree) to 5 (strongly agree). There was no specific training required by raters, beyond the instructions. The BIT team was responsible for survey administration and data entry. Individual mean scores were used for the analysis. This scale is a revised version of the 'General Self-Efficacy Scale', with higher construct validity and reliability than its predecessor (Chen *et al.,* 2001). It has also been found to predict self-efficacy in a number of different domains.
- Teamwork, as assessed by the "Teamwork Scale for Youth" (TSY (Lower-Hoppe *et al.,* 2017). This is an eight-item self-report survey, in which participants mark on a Likert scale from 1 to 5 the accuracy of a series of statements covering examples of positive team behaviour. The scale has good validity, strong reliability in terms of internal consistency, and strong factorial invariance (Lower-Hoppe *et al.,* 2017). The BIT team was responsible for survey administration and data entry. Individual mean scores were used for the analysis.
- Social confidence, as assessed by the "Self-Perceived Communication Competence Scale" (SPCCS (McCroskey and McCroskey, 1988). This is a 12-item self-report survey that measures self-perceived

[8] Our assumption regarding the size of this correlation was based on an analysis of the 'DfE Longitudinal Study of Young People in England' (Anders, 2016, unpublished). This analysis used a linear model of the relationship between Key Stage 2 average (across English, Maths, and Science) points score and the capped 8 points score for GCSE and equivalent qualifications. The dependent and independent variables are slightly different in this current trial (the Key Stage 2 Science assessment has since been dropped and Attainment 8 did not exist previously), but this was the best available analysis that we had to make the estimation.

communication in a range of contexts. Respondents are asked to rate their competence in 12 situations between 1 and 100. This scale has good reliability estimates and performs well in predictive validity tests against similar scales (McCroskey and McCroskey, 1988). It was administered by BIT alongside the Teamwork and Self-efficacy surveys. Individual mean scores are used for the analysis. We also intended to measure part of this construct through the GCSE English Speaking and Listening exam if the data were available for all participants. However, this data is not available via the NPD, so it has not been included.

These three outcomes were chosen, as they directly related to skills that students would be using and developing throughout their time in the programme. A core feature of Community Apprentice is that students work together in a team towards a shared goal, which should enhance their teamwork skills. Further, the students lead on devising, planning, and executing their fundraising activities. Participants also set individual challenge goals of their choosing. Given this self-directed approach, we would expect to see an improvement in self-efficacy. Finally, many fundraising activities require students to collaborate with peers, make requests of school staff, and engage with staff from the business mentor's company—all of which should improve social confidence.

These outcomes were assessed via a single self-report survey, which can be found in **Appendix H**. The survey was administered shortly after pupils had completed the final programme activity (May 2018) by BIT research assistants (RAs). The RAs were given detailed instructions to ensure assessments were delivered consistently. Only those eligible for the intervention and who had agreed to take part had their data collected for secondary outcomes (in both treatment and control arms). Surveys were administered during the time slot allocated to the Community Apprentice sessions, either during a curriculum slot or at a lunch break, depending on the school. The survey was administered on paper and took approximately 15 minutes to complete.[9] The RAs also conducted the data entry for the surveys from their schools. RAs were not blind to assignment because treatment and control surveys were slightly different. The treatment survey included two questions for the IPE on how well pupils felt that they had worked with their Envision coach and business mentors.

Sample size

The sample size was determined by the maximum number of schools and pupils that Envision had capacity to support in the intervention group (30 schools and 13 pupils per school). This gave us a maximum sample size from which we calculated the minimum detectable effect size (MDES) at the protocol stage. Schools were asked to select 26 pupils who were eligible for the intervention, so that we could evenly allocate half to treatment and half to control. In some cases, schools selected less than 26 pupils. In these cases, we randomly assigned 13 pupils to the treatment group (to ensure that all places on the programme were filled), and the remainder to control. This is why there were slightly fewer pupils in the control group (371) as compared to the treatment group (389) post randomisation. More information on the randomisation procedure can be found in the 'Randomisation' section below. There was some pupil-level attrition after randomisation (due to pupil withdrawals) and at the analysis stage (due to missing covariates), see the participant flow diagram in the 'Impact evaluation results' section. No previous studies were identified estimating the effect of a programme like this on academic attainment, so it is hard to assess how well-powered the study was. However, an effect of 0.2 SDs is very large compared to the estimated mean average effect size in the EEF trials, which is 0.04 SD (Demack *et al.,* 2021). The trial may therefore, have been unpowered for the primary analysis and was almost certainly underpowered for the FSM subgroup analysis.

Table 4: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.19	0.34	0.19	0.30	0.19	0.29
Pre- / post-test correlations (r)	Level 1 (pupil)	0.71	0.71	0.71	0.71	0.66	0.60
	Level 2	NA	NA	NA	NA	NA	NA

[9] N.B. Since the survey was administered on paper, randomising items would have been too resource intensive. As such, the items were collected in the same order for all participants.

	(class)						
	Level 3 (school)	NA	NA	NA	NA	NA	NA
Intra-cluster correlations	Level 2 (class)	NA	NA	NA	NA	NA	NA
	Level 3 (school)	NA	NA	NA	NA	NA	NA
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		NA	NA	NA	NA	NA	NA
Number of schools	Intervention	30	30	30	30	30	30
	Control	30	30	30	30	30	30
	Total:	30	30	30	30	30	30
Number of pupils[10]	Intervention	390	117	389	154	345	160
	Control	390	117	371	156	344	161
	Total:	780	234	760	310	689	321

NA, not applicable.

Justification for the assumptions in **Table 4** above are as follows:

- *Pre-/post-test correlations:* In the protocol and randomisation stages, we assumed that 50% of the variance in primary outcome would be explained by Key Stage 2 Maths and Reading scores (for the whole sample and the FSM subgroup). So, the combined R^2 for these baseline covariates was assumed to be 0.5, giving a correlation coefficient of 0.71. This assumption was based on analysis of the 'DfE Longitudinal Study of Young People in England' (Anders *et al.,* 2016, unpublished). At the analysis stage, we adjusted this assumption based on our data. We calculated the R^2 for the baseline covariates by regressing the two covariates on our outcome. We found that the R^2 was actually 0.44 for the whole sample, giving a correlation coefficient of 0.66. R^2 was 0.36 for the FSM subgroup, giving a correlation coefficient of 0.6.
- *Alpha and power:* These are standard assumptions.
- *One- or two-sided test:* A two-sided test was performed to err on the side of caution. There was little existing evidence relating to the effect of interventions of this type on academic attainment, so we could not assume the direction of any effect that we might observe.
- *Number of schools:* The delivery organisation provided the estimate of 30 schools based on their previous experience of recruiting schools to similar interventions and their staff capacity for this intervention. The delivery organisation was able to recruit all 30 schools.
- *Number of pupils:* The delivery organisation provided the estimate of 780 pupils, based on their maximum capacity to support 13 pupils per intervention group. Based on previous experience of running similar interventions, they were confident that schools would commit to this number. In the end, 760 pupils were randomised. However, due to various attrition reasons (see the 'Attrition' section), only 689 were included in the primary analysis.

[10] All figures are for numbers assigned. The MDESs given in this table assumed (for the protocol and randomisation figures) that 23% of pupils would not participate at all in the programme (and so receive no benefit from it) but remain in the sample for analysis. By the analysis stage, 15% of pupils did not participate at all.

- *Free School Meals:* Participating schools were asked to aim to select pupils such that a minimum of 30% were eligible for FSM. We assumed that this aim would be met based on the delivery organisation's prior experience, and because schools were located in areas where intakes had above average proportions of pupils on FSM. Before randomisation, schools shared the FSM data they had on file for the trial participants, which led to 310 pupils eligible for FSM (41% of the total sample) being randomised. For the analysis, we confirmed FSM status using the NPD data, which is more accurate than school records. This led to there being 321 pupils eligible for FSM in the analysis (47% of the total sample).

Randomisation

The study was a two-arm randomised controlled trial with randomisation at the pupil level, stratified by school and FSM status. The two arms were:

1. The treatment arm, in which pupils were assigned a place on the Envision Community Apprentice programme.
2. The control arm, in which pupils continued with routine school activity.

Schools selected pupils for participation in the trial in September 2017 and October 2017. This selection was either done solely by teaching and management staff, or it included a process where pupils put themselves forward for the programme prior to the final selection by staff.[11] Following pupils' initial agreement to take part, there was a two-week opt-out period for pupils and parents/carers. Schools provided lists to BIT of the selected students who did not opt-out of the evaluation.

The total number of pupils who were able to take part in the intervention in each school was limited to 13, as this reflected Envision's capacity to deliver the programme. Pupils were randomised within schools: 13 were randomised into the treatment group; and the remaining 13 assigned to the control group. Where the number of pupils selected exceeded 26, BIT randomly selected 26 pupils to be included in the evaluation, and randomisation was conducted as described above. Where the number of pupils selected was less than 26, 13 were randomised into the treatment group and the remaining number was assigned to the control group. This over-allocation to the treatment group was chosen over even allocation for two reasons: i) the programme model calls for teams of 13 young people and the team wanted to ensure the programme was set-up to run according to the model; and ii) the large team size helps to ensure the work can be fairly distributed across the young people and we were concerned that a small team would struggle to complete the activities, and that this may have led to decreased effectiveness and increased attrition.

The randomisation procedure was conducted using statistical software, Stata Version 14 (StataCorp LLC, College Station, TX, USA). The Stata code used to carry out the randomisation was reviewed, and quality assured by a researcher who was not part of the trial (see **Appendix I**). Randomisation was stratified by school, and also by the Ever 6 FSM variable that indicates whether a pupil has been eligible for FSM in the previous six years. FSM status is likely to be correlated with the primary outcome variable as it is associated with lower attainment on average. Not all schools were able to provide their pupils lists by the original deadline, so randomisation was conducted in three separate rounds so as not to delay communications with schools who had completed recruitment on time. This did not affect random assignment as pupils were randomised within schools.

Statistical analysis

Primary analysis

The primary outcome is overall GCSE performance (Attainment 8). The GCSE Attainment 8 score is calculated as the sum of the points for a pupil's eight best GCSE scores from a set of DfE-approved GCSE subjects, with a double-weighting applied to English and Maths (DfE, 2024). The primary analysis used the ITT framework, in which we tested the hypothesis that being assigned a place on the programme had an effect on attainment. Analysis was carried out using an ordinary least squares (OLS) regression,

[11] No selection criteria were given to staff by Envision for this process, so the decision was at each school's discretion.

$$Y_i = \beta_0 + \beta_1 * T_i + X_i * \alpha + \epsilon_i$$

where:

- Y_i was the outcome for overall GCSE performance, measured by Attainment 8 score;
- T_i was a binary indicator for the treatment assignment (1 if the pupil was assigned to treatment and 0 if not);
- X_i was a vector of individual-level stratification variables (School ID and EVER6_FSM_P) and the baseline attainment measured through separate Key Stage 2 Reading (KS2_READMRK) and Maths (KS2_MATTOTMRK) marks; and
- ε_i was the individual-level error term. Standard errors were corrected for heteroskedasticity.

While point scores were bounded (0 to 90), we assumed that the distribution of the scores was close to normal and response to the treatment was locally linear, so an OLS was appropriate.

Secondary analysis

The secondary analysis was also conducted on an ITT basis, in which we tested the hypothesis that being assigned a place on the programme had an effect on each secondary outcome.

Analysis of the effect on English attainment was carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 * T_i + \beta_2 * X_i + \epsilon_i$$

where:

- Y_i was the outcome for English GCSE performance, measured by the mean average point score (0 to 18)[12] for English Language and English Literature;
- T_i was a binary indicator for the treatment assignment (1 if the pupil was assigned to treatment and 0 if not);
- X_i was a vector of individual-level stratification variables (School ID and EVERFSM_6_P) and the baseline attainment measured through Key Stage 2 Reading mark (KS2_READMRK); and
- ε_i was the individual-level error term. Standard errors were corrected for heteroskedasticity.

Analysis of the effect on Maths attainment was carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 * T_i + \beta_2 * X_i + \epsilon_i$$

where:

- Y_i was the outcome for Maths GCSE performance, measured by point score;
- T_i was a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i was a vector of individual-level stratification variables (School ID and EVERFSM_6_P) and the baseline attainment measured through Key Stage 2 Maths mark (KS2_MATTOTMRK); and
- ε_i was the individual-level error term. Standard errors were corrected for heteroskedasticity.

Analysis of the effect on self-efficacy was carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 * T_i + \beta_2 * X_i + \epsilon_i$$

where:

[12] While Uniform Mark Scale (UMS) scores provide a more granular measure, they are no longer available from the NPD, so point scores are used here.

- Y_i was the outcome for self-efficacy, measured by the mean average score for the individual on the NGSE Scale (0 to 18);
- T_i was a binary indicator for the treatment assignment (1 if the pupil was assigned to treatment and 0 if not);
- X_i was a vector of individual-level stratification variables (School ID and EVERFSM_6_P) and the baseline attainment measured through separate Key Stage 2 Reading and Maths marks (KS2_READMRK and KS2_MATTOTMRK respectively); and
- ε_i was the individual-level error term. Standard errors were corrected for heteroskedasticity.

Academic attainment is known to be correlated with self-efficacy (Gutman and Schoon, 2013), so was included here as a covariate to increase the precision of our estimate.

Analysis of the effect on teamwork was carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 * T_i + \beta_2 * X_i + \epsilon_i$$

where:

- Y_i was the outcome for teamwork, measured by the mean average score for the individual on the TSY;
- T_i was a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i was a vector of individual-level stratification variables (School ID and EVERFSM_6_P); and
- ε_i was the individual-level error term. Standard errors were corrected for heteroskedasticity.

Analysis of the effect on social confidence was carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 * T_i + \beta_2 * X_i + \epsilon_i$$

where:

- Y_i was the outcome for social confidence, measured by the mean average score for the individual on the SPCCS;[13]
- T_i was a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i was a vector of individual-level stratification variables (School ID and EVERFSM_6_P); and
- ε_i was the individual-level error term. Standard errors were corrected for heteroskedasticity.

The evaluation protocol stated that we would also run an ordered logit as a robustness check on the estimates for self-efficacy and teamwork (which are ordinal scales). However, we no longer believe that this is a sensible approach as we are using mean average scores from these scales to calculate the outcome, which effectively makes the outcome variable continuous. An ordered logit robustness check was therefore not included.

Baseline academic attainment has not been used as a control variable in the analysis of teamwork or social confidence as there is no strong evidence, or theoretical argument, to suggest that these outcomes are correlated with attainment.

Analysis in the presence of non-compliance

In the case of one-sided non-compliance (where some individuals assigned to treatment did not participate), we used an instrumental variables approach to estimate the Complier Average Causal Effect (CACE). In the context of the trial, to be considered as minimally compliant with the treatment, a student must have attended at least the following combination of sessions: five weekly sessions; and one business mentoring session.

[13] As specified in the section above on 'Outcome measures', effects were also calculated for the subscores from the SPCCS, which relate to the different forms of communication assessed by the scale: 'Public'; 'Meeting'; 'Group'; 'Dyad'; 'Stranger'; 'Acquaintance'; and 'Friend'.

This minimal dosage was estimated by the delivery organisation (Envision). It is important to note that we do not know the true minimal amount of dosage needed to generate a treatment effect, so the cut-off chosen for minimal compliance was the delivery team's best estimate. This was a qualitative judgement based on the assumptions that: i) attendance at five weekly sessions allows the pupil to progress beyond the introductory sessions into the substance of project goal setting and planning; and ii) interaction with the business mentors was crucial to developing pupils' social confidence. The instrumental variable that we used is treatment assignment, which was assumed to influence participation in the programme but not the outcome variable in its own right.

Two key assumptions need to hold for this approach:

1. Being assigned to the treatment increases participation in the treatment. In this instance, students could only participate in the programme if they were assigned to the treatment group.

2. Random assignment does not in itself impact outcomes. We have no reason to believe that the offer of the programme would improve attainment on its own, but instead believe that any impact on attainment is achieved through participation in the programme.

The CACE estimation used a two-stage least squares (2SLS) approach:

Stage 1

$$Z_i = \beta_0 + \beta_1 * T_i + \beta_2 * X_i + \epsilon_i$$

Stage 2

$$Y_i = \beta_3 + \beta_4 * \hat{Z}_i + \beta_5 * X_i + u_i$$

where:

- Z_i is the binary compliance indicator for individual i;
- T_i is a binary indicator of the treatment assignment for individual i (1 if the pupil is assigned to treatment and 0 if not);
- X_i is a vector of individual-level stratification variables (school and EVERFSM_6_P) and the baseline attainment measured through separate Key Stage 2 Reading (KS2_READMRK) and Maths (KS2_MATTOTMRK) marks for individual i;
- Y_i is the Attainment 8 score for individual i;
- $\hat{Z}_i$ are the predicted levels of compliance from the first stage of the 2SLS;
- ε_i is the error term in the first stage; and
- u_i is the error term in the second stage.

In the protocol we specified an additional descriptive dosage analysis. We did not conduct this analysis for a number of reasons. These are presented in **Appendix J**.

Missing data analysis

First, we reported the number of complete observations (those without any data missing). This identified the following types of missing data:

- missing pre-treatment covariates; and
- missing outcome data.

Data can be missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). These data types and forms of missingness are summarised in **Table 5**. In line with the EEF guidelines on resource allocation, we only conducted imputation as a robustness check for the primary outcome analysis. Schultz and Grimes (2002)

suggest that, when less than 5% of data is missing, there is likely to be little bias introduced to estimated treatment effects. We have adopted this threshold here, specifying imputation for covariates where more than 5% of the data is missing.[14]

Table 5: Types of missing data

Type of missing variable	Type of missing data	Further analysis
Pre-treatment covariates	MAR	Multiple imputation (MI) before analysis of treatment effects
Pre-treatment covariates	MNAR	Sensitivity analysis
Outcome data	NA	None. Missing outcome data will not be imputed, and these observations will therefore be lost to analysis

NA, not applicable.

Missing pre-treatment covariates

All observations with missing pre-treatment covariates were included in the analysis if the outcome measure and treatment assignment were not missing. We had complete data for School ID (identification) and EVER6_FSM_P. However, 7% of the Key Stage 2 attainment data was missing (for both Maths and Reading). This was due, for example, to pupils in the sample who did not sit the test(s). To account for this, we first try to establish, which variables are predictive of the missing data. To do this, we created a new variable that is a binary indicator of missingness and looked for its predictors using a logistic regression model. Missing Key Stage 2 attainment data was modelled as follows:

$$M_i \sim binomial(p_i); logit(p_i) = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_4 * T_i$$

where:

- M_i was the binary variable for missingness (equal to 1 if missing and 0 if not missing);
- pi was the probability that a given observation is missing the Key Stage 2 Reading mark;
- X_1 was the EVER6_FSM_P variable;
- X_2 was the Key Stage 2 Maths or Reading mark;
- X_3 was a vector of binary variables for the School ID;[15] and
- T_i was the treatment assignment indicator (equal to 1 if assigned to treatment and 0 if assigned to control).[16]

We then followed the following rules for how to model the missing data:

- If the coefficients in the regression are significant (i.e. the values were missing conditional upon other variables in the model) and missingness does not depend on unobserved covariates, imputation will provide an unbiased estimate of the true values. Multiple imputation (MI) should be carried out using the Markov chain Monte Carlo (MCMC) method to predict the missing values prior to the analysis of treatment effects. We will then estimate the treatment effect using the imputed data and compare our result with the primary analysis (conducted on complete cases only). If the point estimates of these two datasets are not similar (more than 0.05 standard deviations apart), it is likely that the data is MNAR and sensitivity analysis will be carried out. This will entail modelling missingness as above with all possible combinations of the variables available (EVER6_FSM_P, Key Stage 2 Reading and Key Stage 2 Maths scores, and School ID).

[14] This is also in line with the convention provided in the EEF guidance on statistical analysis.
[15] N.B. As School ID is a categorical variable with 30 categories, an F-Test will be used to test for the overall significance of school ID.
[16] The indicator for treatment assignment was missed out of this specification in the Statistical Analysis Plan. This was a typo so is included here and in the analysis. See **Appendix J**.

- If, after modelling missingness, as described above, it is found that our covariates do not explain the missingness, this will imply that the data is either MCAR or MNAR. In this case, we will be conservative and assume that the data is MNAR and conduct sensitivity analysis. These sensitivity analyses will investigate the sensitivity of the point estimate of the treatment effect to changes in model specification (and hence, sample definition), through the inclusion and exclusion of variables for which observations are missing, as well as using null imputation to provide a more intuitive analysis based on a full sample of data.

The results of this modelling exercise, and how we applied these rules, is described in the 'Missing data analysis' section below.

Missing outcome data

No intermediate outcomes were identified as appropriate to use to impute primary outcome data; the secondary outcomes in this study (teamwork, communication confidence, and self-efficacy) are not sufficiently predictive of the primary outcome. Observations with missing primary outcome data were therefore, dropped from the primary analysis and a complete case analysis was run.

Subgroup analyses

We estimated the effect on the primary outcome for the subgroup of pupils who were registered for FSM in the NPD (using the EVERFSM_6_P variable). We used the same model as our primary analysis, with the addition of an interaction between treatment assignment and FSM status, to assess whether there is a significant difference in the treatment effect between FSM students and others. The model used for this analysis was as follows:

$$Y_i = \beta_0 + \beta_1 * T_i + \beta_2 * X_i + \beta_3 * Z_i + \beta_4 * T_i * Z_i + \epsilon_i$$

where:

- Y_i was the outcome for overall GCSE performance, measured by Attainment 8 score;
- T_i was a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i was a vector of the other individual-level stratification variables (School ID) and the baseline attainment measured through separate Key Stage 2 Reading (KS2_READMRK) and Maths (KS2_MATTOTMRK) marks;
- Z_i was a binary variable indicating the FSM status (EVERFSM_6_P) of individual i; and
- ε_i was the individual-level error term. Standard errors were corrected for heteroskedasticity.

In this model, β_1 denotes the effect of the invention for participants who were not eligible for FSM in the prior six years ($Z_i = 0$) and $\beta_1 + \beta_4$ denotes the effect of the intervention on those who were eligible for FSM in the prior six years ($Z_i = 1$). The difference in intervention effects between the two groups is therefore denoted by β_4.

As per the EEF guidelines, we also estimated the treatment effect for the subsample of participants who were eligible for FSM in the prior six years (EVERFSM_6_P = 1) and compared this to the estimated treatment effect for those not eligible for FSM (EVERFSM_6_P = 0). This was operationalised using the regression model in the primary analysis for each of these two groups.

Additional analyses and robustness checks

Imbalance at baseline

We assessed imbalance at baseline, and for the subsample of those analysed, by calculating the following values in each case and cross-tabulating by treatment arm:

- count and percentage of male participants;
- count and percentage of female participants;
- mean and SD of Key Stage 2 Maths mark;

- mean and SD of Key Stage 2 Reading mark; and
- percentage of Ever 6 FSM.

The difference in mean Key Stage 2 scores between the treatment and control group are expressed as Hedges' g effect sizes.

Estimation of effect sizes

Effect sizes are expressed in terms of Hedges' g, using the following formula:

$$ES = \frac{M_1 - M_2}{SD^*_{pooled}}$$

where,

- M_1 is the mean value of the outcome in the control group; and
- M_2 is the mean value of the outcome in the treatment group.

and,

$$SD^*_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

where,

- n_1 is the number of observations analysed in the control group;
- n_2 is the number of observations analysed in the treatment group;
- SD_1 is the SD of the outcome variable in the control group; and
- SD_2 is the SD of the outcome variable in the treatment group.

The difference in means for each outcome variable were recovered from the relevant regression (β_1).

Longitudinal analysis

Four longitudinal outcomes may be examined where it is possible to access the relevant data. These are:

- post-16 attainment (Key Stage 5) sourced through the NPD;
- progression into further social action, sourced via the NCS;
- progression to higher education, sourced via the Longitudinal Education Outcomes (LEO) dataset; and
- employment status sourced through the LEO dataset.

At the time of publishing the trial protocol and Statistical Analysis Plan, we anticipated reporting on Key Stage 5 and social action outcomes in 2022 and higher education and employment outcomes in 2023. However, substantial delays in accessing the NPD data for the primary analysis mean that we are reporting the main trial results in 2024. The EEF's approval of the longitudinal analysis is conditional on the results found in this main trial, so are not included in this report.

Implementation and process evaluation

Methods

A mixed-methods approach was taken to data collection, combining evidence from interviews, observations, surveys, and administrative data. The approach was 'exploratory sequential', so that findings from each stage were used to inform the approach to subsequent stages (Teddie and Tashakkori, 2009, p. 120). For example, themes that emerged from early case study interviews and observations were used to refine the design of interview and observation guides for later

visits. A case study approach was taken to the qualitative part of the study. Three schools were selected for case study following the sampling strategy described below. In each school, we observed an intervention activity, and interviewed pupils, school staff, and business mentors, allowing us to explore these different perspectives within the context of a single school's adoption of the programme (Ritchie *et al.,* 2014, p. 66). Five pupils were interviewed in each school and the samples of school staff and mentors varied by case, as described in more detail below.

Observations

One intervention activity was observed in each case study school by BIT researchers. The timing and type of activity observed varied by school. The business mentor sessions were thought by the developers to be particularly important components of the intervention, so two of these sessions were observed in the sponsoring businesses offices; with one school near the beginning of the programme, in the planning phase, and with another school towards the end of the programme, when pupils were reflecting on their achievements and preparing their pitch for the boardroom final. In the third school, a weekly in-school session, led by the Envision coach, was observed towards the end of the programme. This session also supported pupils to prepare their pitch for the boardroom final. This variation in sessions allowed us to observe the intervention being delivered in different environments (business office vs school), with different facilitators (business mentors vs Envision coach), and at different times during the programme (planning phase vs reflection phase). These observations served two purposes: to familiarise the researchers with the intervention (supporting better depth in interviews); and to provide behavioural information to complement interview data. They focused on five themes: pupil comprehension; pupil engagement; the relationship between pupils and facilitator(s); the physical environment; and potential causal mechanisms. Semi-structured observation frameworks were used in all cases.

Interviews

In all case study schools, interviews were carried out with pupils and a selection of staff and volunteers involved in the delivery or management of the intervention, namely, a combination of a business mentor, the contact teacher, and the relevant senior manager at the school. These interviews were semi-structured and focused on fidelity, adaptation, quality, reach, and responsiveness of pupils to the programme. A first round of interviews was conducted in all case study schools directly after the observations described above. Researchers also returned to two schools after the boardroom final - one that won the competition in their region and one that lost—to interview the same pupils again, to explore the potential moderating effects of winning and losing. **Table 6** below, summarises the qualitative data collection in the case study settings.

Table 6: Data collection in case study settings[17]

Case study setting	Session observed	Timing of observation	Roles interviewed (n)	Post-intervention interviews?
School 1	Business mentoring	February 2018	Business mentor (n=1); senior manager (n=1); pupils (n=5)	No
School 2	Weekly session	March 2018	Contact teacher (n=1); senior manager (n=1); pupils (n=5)	Yes
School 3	Business mentoring	April 2018	Business mentor (n=1); senior manager (n=1); contact teacher (n=1); pupils (n=5)	Yes

Administrative data

Student registration data was used to establish the proportion of intervention pupils that were entitled to FSM in the last six years; an indicator that the intervention was reaching its intended participants. Session registers were used to establish intervention pupil attendance; an indicator of engagement and compliance (see CACE analysis in 'Analysis in the presence of non-compliance' section below for more details).

[17] The protocol specified that at least one business mentor, senior manager, and pupil would be interviewed in every case study school. After the first case study visit, however, it was agreed with the developers that we would replace a business mentor for a contact teacher in one setting. The rationale for this change was that, of all staff and volunteers involved in the delivery and management of the intervention (other than Envision staff), the contact teacher seemed to have the most detailed knowledge of the intervention's implementation and of pupil development.

Surveys

To assess whether pupils were engaged in activities that were similar to the intervention, all members of both the intervention and control groups were asked in an endpoint survey to report whether they participated in any extracurricular activities during the trial period. They were also asked to give a description of these activities. To provide a simple quantitative indicator of the quality of the relationships between pupils and their mentors and coaches, the endpoint survey for the intervention group also asked participants to rate these two relationships on a scale of 1 to 5.[18]

Table 7: IPE methods overview

Research methods	Data collection methods	Participants / data sources (n)	Data analysis methods	Research questions addressed	Implementation / logic model relevance
Case studies	Semi-structured interviews	Pupils (n=15) Business Mentors (n=2) Contact teachers (n=2) Senior managers (n=3)	Deductive coding; inductive coding; thematic analysis	1.a; 1.b; 1.c.i; 2.a.ii; 2.a.iii; 2.b.ii; 2.b.i; 2.b.iii; 2.b.iv	Fidelity; dosage; adaptation; quality; programme differentiation; reach; responsiveness; logic model
Observations	Semi-structured observations	Intervention sessions (n=3)	Thematic analysis	1.a; 2.b.ii	Fidelity; dosage; responsiveness; logic model
Surveys	Paper questionnaires	Pupils (n=686)	Descriptive statistics; regression	1.b; 1.c.i; 1.c.ii	Quality; programme differentiation
Administrative data	National Pupil Database (for FSM); Pupil registers (for attendance)	Pupils (n=760)	Descriptive statistics (for FSM); regression (for CACE analysis)	2.a.ii; 2.b.i	Reach; responsiveness

Sampling

For the qualitative elements of the study, sampling was purposive, with units selected for variation on characteristics that were thought to be particularly relevant to the research questions. Three schools in the treatment group were selected for case studying on this basis. Sampling of these schools was carried out by the researchers, aiming for variation in the following characteristics.[19]

- location;
- type of business providing mentors;
- engagement with the intervention (indicated by the 'buy-in' of the school's senior management—i.e. their willingness to actively support and participate in the programme—qualitatively assessed by Envision);[20] and
- Office for Standards in Education, Children's Services and Skills (Ofsted) rating.

Within those schools, the contact teacher selected pupils for interview with the aim of getting variation in levels of engagement with the intervention, based on their own judgement. Business mentors were convenience sampled based on their availability on the day of the intervention observations, and the contact teachers and senior managers interviewed were those responsible for implementation and oversight of the intervention in their school, respectively.

The purposive approach to sampling has important implications for the analysis and findings. The aim of this sampling method is to capture the range and diversity of experiences in relation to the research questions. Importantly, this approach is not intended to generate a sample that is statistically representative of either the study population or the

[18] The methods map in the protocol suggested that the endpoint survey would also be used to partially identify barriers and facilitators to participation. Once the survey was drafted however, it was considered too burdensome to include questions on these topics on top of the secondary outcome scales and questions about extracurricular activities. In any case, these topics are more effectively explored through qualitative interviews.
[19] The protocol did not specify whether these criteria should be ordered but, following discussions with the developers, they were prioritised in the order listed. As such, less variation was achieved by Ofsted rating than the other criteria.
[20] The protocol suggested that engagement would be assessed by reviewing a range of administrative data. However, no such data was available from the developers at the point of sampling, so they instead provided a qualitative assessment of senior management buy-in based on their interactions to date.

wider population from which the total study sample was drawn. As such, reporting the prevalence of an experience in the qualitative findings 'tells us nothing about the prevalence within [either] population' (Ritchie *et al.,* 2014, p. 329). Furthermore, qualitative methods, by their nature, do not collect data in the structured way that is necessary for quantitative aggregation. This is in contrast, for example, to structured survey questionnaires, which collect responses in fixed categories that can be aggregated. The reporting of frequency counts in relation to qualitative findings is therefore, carefully avoided as such counts are at best uninformative and at worst misleading.

Table 8: Characteristics of case study schools

Case study setting	Location	Business type	Senior management buy-in	Ofsted rating
School 1	Bristol	Military services	High	Outstanding
School 2	Birmingham	Events and entertainment	Low	Good
School 3	London	Credit assessment	High	Outstanding

Analysis

Three analyses were carried out for the IPE: analysis of the administrative data; analysis of the qualitative case study data; and analysis of the survey data.

Administrative data analysis

From the administrative data, a percentage was calculated for the proportion of pupils eligible for FSM. Analysis of the pupil attendance data is described in the section 'Analysis in the presence of non-compliance' on CACE analysis below.

Qualitative data analysis

A thematic analysis was carried out on the qualitative data. This analysis included a three-stage process of organising and analysing the data from interview transcripts. In the first stage, transcripts were coded by research question for the purpose of 'topic coding' (Saldana, 2009). In the second stage, they were coded by theme in response to the relevant research question. These themes were identified inductively from the data (Braun and Clarke, 2006, p. 83). In the third stage, these themes were refined by reviewing their relation to each other and to the data collected from observations and surveys, and to ensure that they comprehensively covered the data (including all perspectives on the research questions) (Ritchie *et al.,* 2014, p. 280). In this final stage of analysis, searches were carried out for thematic categories and patterns across different cases, linkage between phenomena within one case, and linkage in phenomena between groups of cases (though the latter was limited by the small number of cases) (Ritchie *et al.,* 2014, p. 280).

Survey data analysis

To help assess whether pupils in the trial were engaged in activities that were similar to the intervention, participants in both the intervention and control groups were asked in the endpoint survey to report whether they participated in any extracurricular activities. They were also asked to give a description of these activities. Analysis of these survey items proceeded in two steps. First, the descriptions were coded by the researcher into one or more of three categories: i) activities that could have a positive effect on the primary outcome; ii) activities that could have a positive effect on a secondary outcome; and iii) youth social action activities. An activity was put in category (i) if it was listed in the EEF Toolkit as having a positive effect. An activity was put in category (ii) if the researcher judged there to be a plausible theory of change that linked it to one or more of the secondary outcomes. For example, all team sports were placed in category (ii) as they would plausibly improve participants' teamwork skills, social confidence, and general self-efficacy. A debating club would also be placed in category (ii) because it could improve an individual's social confidence and general self-efficacy. An activity was classified as 'youth social action' in category (iii) if it met the industry standard definition; programmes that support young people to take 'practical action in the service of others' (Pye and Michelmore, 2017). Examples of reported activities that fell into this category include uniformed youth groups and volunteering with local religious groups. This coding process was necessarily imprecise as survey respondents gave very brief descriptions of their activities (e.g. 'football club'). An inclusive approach was therefore, taken to coding, where an activity was included in a category if it was plausible that it could meet the relevant criterion.

After this process of coding was complete, the following comparisons were made between the intervention and control groups:

- percentage of respondents that reported doing extracurricular activities;

- percentage of respondents that reported doing extracurricular activities that could cause a positive effect on the primary outcome;
- percentage of respondents that reported doing extracurricular activities that could cause a positive effect on a secondary outcome; and
- percentage of respondents that reported doing youth social action (other than the intervention) as part of their extracurricular activities.

Confidence intervals (CIs) were then estimated on the differences between groups using logistic regressions of the following form.

$$Y_i \sim binomial(p_i); logit(p_i) = \beta_0 + \beta_1 * T_1$$

where:

- Y_i is binary indicator of category membership (equal to 1 if the activity is in the category and 0 if not);
- p_i is the probability that a given participant does the activity in question; and
- T_i is the treatment assignment indicator (equal to 1 if assigned to treatment and 0 if assigned to control).

The survey questions for this analysis were included in the trial protocol, but the details of this analysis were not pre-specified. As such, the regression model was kept to its simplest form, without covariates.

Costs

Information on the cost of the intervention was collected from two sources. Envision contributed the amount that it received from the EEF to subsidise delivery of the intervention, and the amount of additional subsidies sourced from corporate sponsors. Participating schools provided estimates of the costs that they incurred as a result of participation in the intervention. For the latter category, structured interviews were conducted with 26 of the 30 schools participating in the intervention (see **Appendix K** for a copy of the interview guide used).[21] These interviews were carried out during the visits to each school for secondary outcome data collection or as follow-up calls if staff were not available during data collection. The four remaining schools were unavailable for interview during the data collection period. For each type of cost identified at the school level, a mean average was calculated. The following approach was taken to calculating the cost per pupil per year.

1. The average cumulative cost incurred by a school over three years was estimated (where 'year one' is the trial year). The costs of years two and three were assumed to be equal to the cost of year one, minus any one-off start-up costs.

2. The cumulative cost over three years of the EEF subsidy was calculated (assumed to be the same in each year).

3. The sum of the costs calculated (in bullet points 1 and 2 above) was divided by three, to give an average cost per school per year.

4. The average cost per school per year was then divided by 13 (i.e. the number of places available for pupils per school).

[21] The protocol specified that we would collect preliminary cost information from schools during case study visits for the IPE, with the possibility of following up with a survey for teachers. After discussion with the developers, it was decided that there was likely to be a wide range and diversity of costs incurred by schools. Interviews in case study schools and a cost survey (which would not give the opportunity to probe responses) were therefore, considered to be insufficient methods, so structured interviews were attempted in all schools.

As the intervention was further subsidised by Envision, through additional corporate sponsorship, figures were also calculated for the following two quantities:

- the average unsubsidised cost per pupil year; and
- the projected future actual cost to schools per year (taking into account any ongoing subsidy that Envision intends to provide).

Timeline

Table 9: Timeline

Dates	Activity	Responsible
October 2017 – November 2017	Schools selected eligible pupils for trial and confirmed student interest in participation	Envision
By end November 2017	Opt-out forms were provided to students and parents/carers	Envision
w/c 4th December 2017	School sent BIT student data for those who did not opt-out	Envision
w/c 11th December 2017	Students were randomly allocated to two groups by BIT	BIT
9th January 2018 17th January 2018	Intervention began in Birmingham Intervention began in Bristol and London	Envision
2nd January 2018	Case study data collection began	BIT
27th April 2018 4th May 2018	Intervention ended in Birmingham Intervention ended in Bristol and London	Envision
30th April 2018 – 15th June 2018	Secondary outcome data and participant IPE data collected	BIT
20th July 2018	Case study data collection ended	BIT
2nd July 2019	Students sat GCSEs	NA
April 2023[22]	BIT collected primary outcome data from the NPD	BIT
27th March 2024	BIT submitted draft evaluation report to the EEF	BIT
15th August 2024	The EEF peer review and feedback completed	BIT
September 2024	The EEF and BIT published the report	BIT

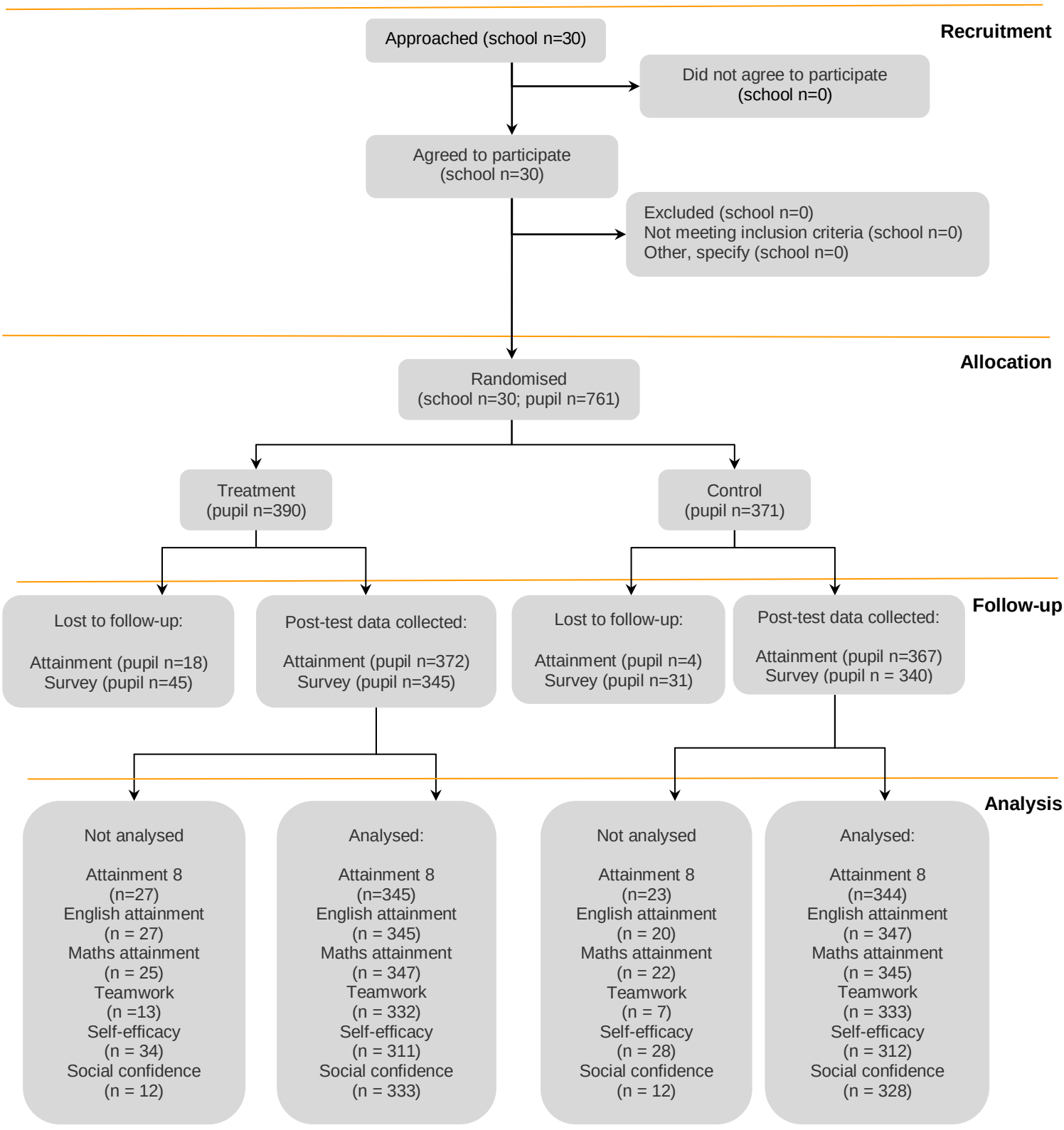
NA, not applicable; w/c, week commencing.

[22] This delay was primarily due to the shift in the method for accessing the NPD. There is no way that this could impact the results, since the data and outcomes will not have been impacted by the delays.

Impact evaluation

Participant flow including losses and exclusions

Figure 2: Participant flow diagram (two arms)



The MDES estimated at various points of the trial is reported in **Table 4**. For our primary outcome, this was 0.19 SD at the protocol design and randomisation stages. This stayed consistent at the analysis stage. As pupils with FSM are a subset of all pupils, the sample size for FSM is smaller and power is substantially lower, at 0.29 at the analysis stage.

Attrition

The number of pupils analysed refers to those with both outcome and complete covariate data. Given the attainment data and the other secondary outcomes come from different sources, the rate of attrition varies. Additionally, even within a data source the attrition rate varies. This is because the covariates we use in the regressions vary by outcome, and as such, the number of students included in the analysis varies by outcome. Below we present the attrition rate across both sources, as well as for each outcome individually.

Attainment data

All attainment data came from the NPD. Reasons for attrition from the attainment outcomes were recorded for all pupils. A total of 11 students in the treatment group withdrew from the evaluation. When matching our data with the NPD, an additional 11 students were not matched. For seven students, this was because they completed Key Stage 4 in the following school year, not in our year of interest. For four students, they did not have any Key Stage 4 data. The remaining students did not have the relevant Key Stage 2 data available. We do not have the reason for missing Key Stage 2 data at the pupil level. However, the DfE reports that Key Stage 2 data can be missing for pupils for the following reasons: they were in an independent school for Key Stage 2; they were not in a school in England; or they have a new UPN in Key Stage 4 for some reason.

Table 10: Pupil-level attrition from the trial (primary outcome, Attainment 8)

		Intervention	Control	Total
Number of pupils	Randomised	390	371	761
	Analysed	345	344	689
Pupil attrition (from randomisation to analysis)	Number	45	27	72
	Percentage	12%	7%	9%

Table 11: Pupil-level attrition from the trial (secondary outcome, Maths attainment)

		Intervention	Control	Total
Number of pupils	Randomised	390	371	761
	Analysed	345	347	692
Pupil attrition (from randomisation to analysis)	Number	45	24	69
	Percentage	11%	6%	9%

Table 12: Pupil-level attrition from the trial (secondary outcome, English attainment)

		Intervention	Control	Total
Number of pupils	Randomised	390	371	761
	Analysed	347	345	692
Pupil attrition (from randomisation to analysis)	Number	43	26	69
	Percentage	11%	7%	9%

Survey data

Reasons for attrition from the measurement of secondary outcomes were recorded for all pupils. Overall, we had 685 pupils complete the survey. A total of 30 pupils opted out of taking the survey, 29 were absent on the day of the survey and at mop-up sessions, and 17 pupils had left the school. Mop-up sessions were attempted at all schools where more than two students were absent on the day of survey administration. It is worth noting that attrition rates for the survey outcomes vary significantly. This is because for every question that someone did not answer, the relevant outcome was missing, and some questions were more commonly skipped than others.

Table 13: Pupil-level attrition from the trial (secondary outcome, teamwork)

		Intervention	Control	Total
Number of pupils	Randomised	390	371	760
	Analysed	332	333	665
Pupil attrition (from randomisation to analysis)	Number	58	38	76
	Percentage	15%	10%	10%

Table 14: Pupil-level attrition from the trial (secondary outcome, self-efficacy)

		Intervention	Control	Total
Number of pupils	Randomised	390	371	761
	Analysed	311	312	685
Pupil attrition (from randomisation to analysis)	Number	79	59	76
	Percentage	21%	16%	10%

Table 15: Pupil-level attrition from the trial (secondary outcome, social confidence)

		Intervention	Control	Total
Number of pupils	Randomised	390	371	760
	Analysed	333	328	685
Pupil attrition (from randomisation to analysis)	Number	57	43	76
	Percentage	15%	12%	10%

Pupil and school characteristics

Table 16 summarises the baseline pupil-level characteristics of intervention and control pupils as randomised. In general, it shows that intervention and control pupils were similar to each other, but both differed in certain ways compared to the national-level figures.

The intervention group had a slightly lower proportion of FSM pupils (44%) versus the control group (47%). Compared to the national average of 15%, the sample had a much higher proportion of FSM pupils, which was achieved through targeted school recruitment. The two groups had a similar gender split although the intervention group had slightly more

males than the control group (intervention 45% vs control 43%). Compared to the national average of 51%, the sample had a lower proportion of males.

The difference in Key Stage 2 scores between the two groups are minimal. The control group has marginally higher average Maths and Reading scores. For Maths, the difference comes to 1.79 out of 100, or a Hedges' g of 0.074. For Reading, the difference in scores is 0.65 out of 100, or a Hedges' g of 0.078.

Table 16: Baseline characteristics of pupils as randomised (excluding withdrawals)[23]

	National-level mean[24]	Intervention group		Control group	
Pupil level (categorical)		n/N[25] (missing)	Count (%)	n/N (missing)	Count (%)
FSM	15%	166/379 (7)	166 (44)	174/371 (4)	174 (47)
Female	49%	207/379 (0)	207 (55)	210/371 (0)	210 (57)
Male	51%	172/379 (0)	172 (45)	161/371 (0)	161 (43)
Pupil level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
Key Stage 2 Maths	Not available[26]	347/379 (32)	71.87 (19.16)	345/371 (26)	73.26 (18.31)
Key Stage 2 Reading	Not available	345/379 (34)	30.76 (8.12)	347/371 (24)	31.41 (8.54)

Table 17 presents the analogous balance characteristics for the groups as analysed for the primary outcome (Attainment 8). As analysed, the very small differences between the control and intervention group we observed as randomised ultimately disappear and the distributions become almost identical. All analyses control for FSM and Key Stage 2 scores.

Table 17: Baseline characteristics of pupils as analysed[27]

	National-level mean[28]	Intervention group		Control group	
Pupil level (categorical)		n/N (missing)	Count (%)	n/N (missing)	Count (%)
FSM	15%	161/345 (0)	161 (47%)	160/344 (0)	160 (47%)
Female	49%	194/345 (0)	194 (56%)	194/344 (0)	194 (56%)
Male	51%	151/345 (0)	151 (44%)	150/344 (0)	150 (44%)
Pupil level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
Key Stage 2 Maths		345/345 (0)	72.13 (18.91)	344/344 (0)	73.41 (18.11)
Key Stage 2 Reading		345/345 (0)	30.76 (8.12)	344/344 (0)	31.42 (8.53)

[23] N.B. That percentages are calculated excluding missing data, showing us the percentage of those with data. For example, for FSM, we report that 43% of those that we have FSM data for are FSM students.
[24] https://explore-education-statistics.service.gov.uk/data-tables.
[25] N.B. That the total N does not match that presented in the participant flow diagram. This is because we do not have the baseline characteristics of the 11 pupils who withdrew from the trial.
[26] Only scaled scores for national averages are published by the DfE. These are not comparable to the raw marks used in our analysis.
[27] N.B. That percentages are calculated excluding missing data, showing us the percentage of those with data. For example, for FSM, we report that 43% of those that we have FSM data for are FSM students.
[28] https://explore-education-statistics.service.gov.uk/data-tables.

Table 18 summarises the baseline school-level characteristics and how they compare to national-level figures. Around 100% of schools in the sample were urban, compared to 75% of schools at the national level. There were 5 community schools, 18 academies, and 7 other types of schools. Four schools were rated by Ofsted as outstanding, 20 schools rated as good, 3 schools rated as requiring improvement, and 1 school rated as serious weakness. In sum, compared to the national average, the analysis sample is more urban, has almost double the percentage of academy schools, and has schools that perform slightly higher than the national average, but is broadly similar in Ofsted rating.

The location, school type, and school performance could moderate the effects of the intervention. The effects identified in our analysis may therefore be different if the treatment were scaled up to all schools in England. The IPE findings suggest that the quality of the Envision coach is an important factor in the success of the programme. Recruiting high-quality coaches is probably easier in urban areas where the population density is higher. So, if the programme were scaled up nationally, an increase in the proportion of rural schools in the sample may lead to a reduction in the effects.

Table 18: Baseline characteristics of schools as randomised[29]

School level (categorical)[30]		National-level proportion[31]	N (missing)	Count (%)
Location[32]	Urban	75%	28 (2)	28 (100%)
	Rural	24%	0 (2)	0 (0%)
School type	Community	26%	5 (0)	5 (17%)
	Academy	34%	18 (0)	18 (60%)
	Other	40%	7 (0)	7 (23%)
Ofsted rating	Outstanding	15%	4 (2)	4 (14%)
	Good	72%	20 (2)	20 (71%)
	Requires improvement	10%	3 (2)	3 (11%)
	Serious weaknesses	1%	1 (2)	1 (4%)
School level (continuous)		National-level mean[33]	N (missing)	Mean (SD)
Key Stage 4 average overall performance		40.07	30 (0)	47.04 (8.17)
Key Stage 4 average English performance		8.35	30 (0)	10.09 (1.50)
Key Stage 4 average Maths performance		7.57	30 (0)	9.04 (1.79)

Outcomes and analysis

Primary analysis

The primary outcome measure was the Attainment 8 score, a measure of overall GCSE performance. It is the sum of the points for a pupil's eight best GCSE scores from a set of DfE-approved GCSE subjects, with a double-weighting applied to English and Maths. Attainment 8 scores can take a value from 0 to 90, so the estimated effect can take theoretical values between -90 and 90. Effects are also presented as Hedges' g to make it easier to compare between outcomes and with other studies.

Figure 3 shows the distribution of outcomes.[34] Scores are distributed across the possible range with no observable skew; as such, we do not have any concerns about floor or ceiling effects in the analysis.

[29] N.B. Percentages are calculated excluding missing data, showing us the percentage of those with data.
[30] https://explore-education-statistics.service.gov.uk/data-tables.
[31] https://explore-education-statistics.service.gov.uk/data-tables.
[32] N.B. This data was not available for the school year 2018/2019. As a proxy, we use 2019/2020 data with the assumption that not many schools change location.
[33] https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-4-performance.
[34] N.B. This graph excludes any counts below ten to be compliant with the Office for National Statistics (ONS) policies.

Figure 3: Histogram of raw Attainment 8 scores

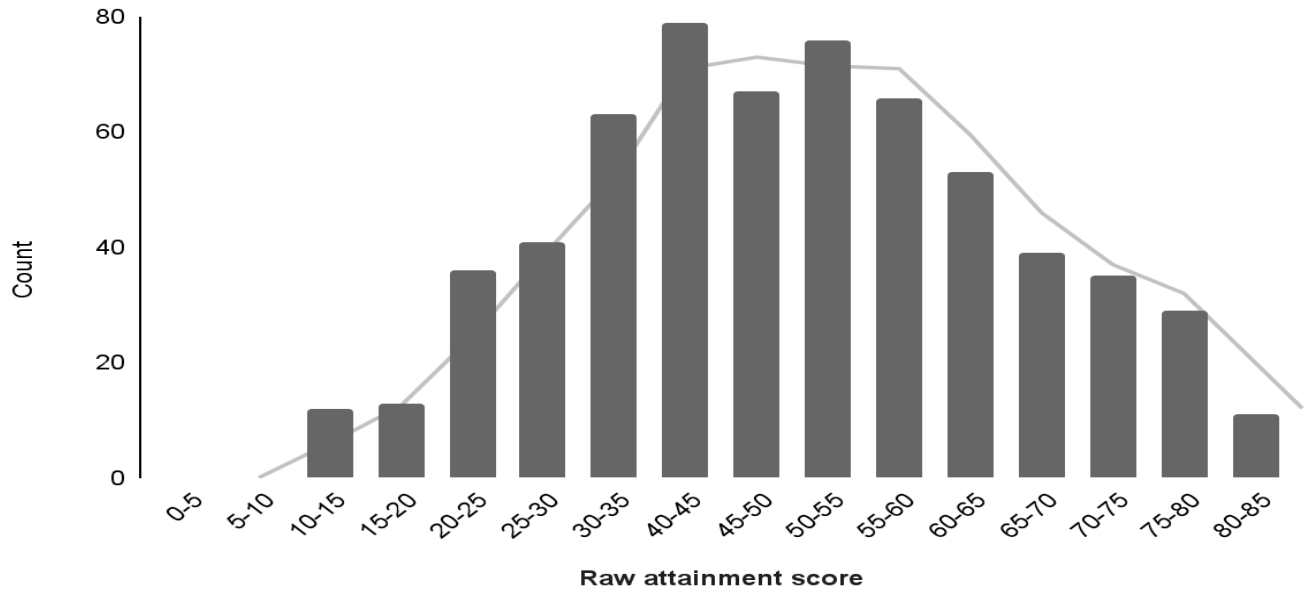


Table 19 presents the results of the analysis for the Attainment 8 outcome (see 'Primary analysis' section above for details of the analysis specification). The unadjusted mean for the Attainment 8 score in the intervention group is 49.04 and 49.91 in the control group (out of 90). After adjusting for covariates in the analysis model, the mean difference between the two groups is 0.12 points, which translates into a Hedges' g effect size of 0.0067 (equivalent to zero month's additional progress). This is our best estimate of the effect but, at the 95% confidence level, the results are compatible with effects that range from substantially negative (two months' less progress) to substantially positive (two months' additional progress). There is therefore, a very large amount of uncertainty in the estimate. The very wide confidence interval (CI) is a result of the study only being powered to detect a large effect due to the small sample size.

Table 19: Primary analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention, control)	Hedges' g (95% CI)	P-value
Attainment 8 score	345 (45)	49.04 (47.11, 50.96)	34 (27)	49.91 (47.94, 51.88)	689 (345, 344)	0.0067 (-0.096, 0.11)	0.90

Secondary analysis

Attainment

The two secondary attainment outcome measures were GCSE English score, a measure of English attainment and GCSE Maths point, a measure of Maths attainment. Both scores can take a value from 0 to 18, so the estimated effect can take theoretical values between -18 and 18. Effects are also presented as Hedges' g to make it easier to compare between outcomes and with other studies.

Figure 4 and **Figure 5** show the distributions of outcomes.[35] Scores are distributed across the possible range with no observable skew; as such, there are no concerns about floor or ceiling effects in the analysis.

[35] N.B. This graph excludes any counts below ten to be compliant with ONS policies.

Figure 4: Histogram of raw Maths attainment scores

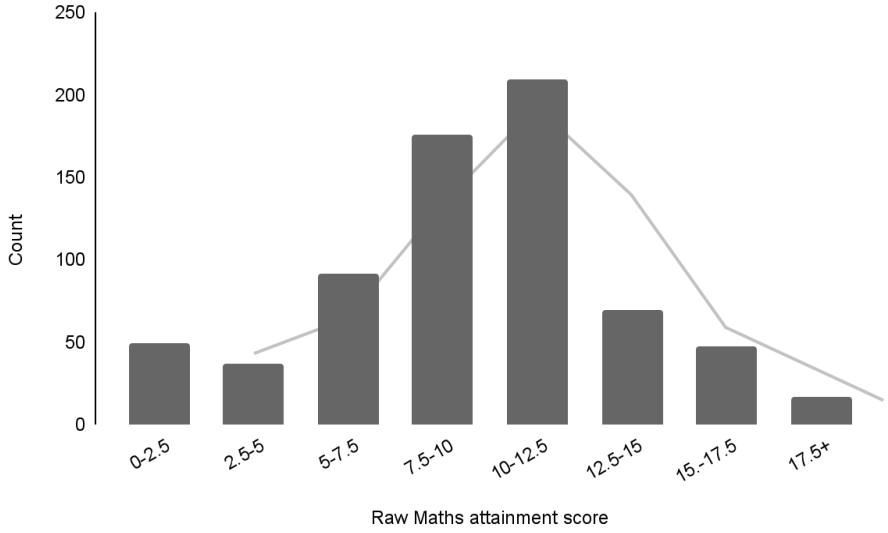


Figure 5: Histogram of raw English attainment scores

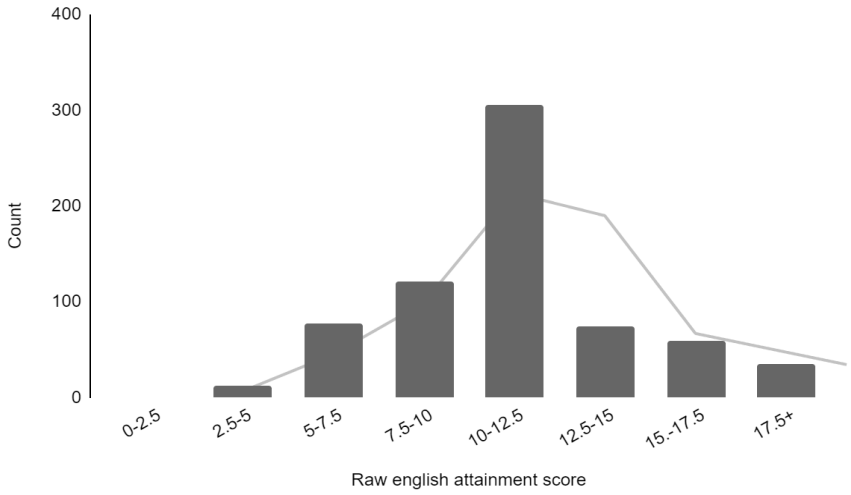


Table 20 presents the results of the analysis for the two secondary attainment outcomes (see 'Secondary analysis' section above for details of the analysis specification), the findings of which are in line with the primary analysis.

For the English score, the unadjusted mean in the intervention group is 10.49 and in the control group is 10.73 (out of 18). After adjusting for covariates in the analysis model, the mean difference between the two groups is -0.073, which translates into a Hedges' g effect size of -0.020 (a negligible negative effect, or zero month's progress). This is our best estimate of the effect but, at the 95% confidence level, the results are compatible with effects that range from substantially negative (two months' less progress) to substantially positive (two months' additional progress). There is therefore, a very large amount of uncertainty in the estimate. The very wide CI is a result of the study being powered to detect a large effect.

For the Maths score, the unadjusted mean in the intervention group is 9.35 and that in the control group is 9.61 (out of a potential total score of 18). After adjusting for covariates in the analysis model, the mean difference between the two groups is -0.085. The Hedges' g effect size for the difference between groups is -0.021 (a negligible negative effect, or zero month's progress). This is our best estimate of the effect but, at the 95% confidence level, the results are compatible with effects that range from substantially negative (two months' less progress) to substantially positive (two months' additional progress). There is therefore, a very large amount of uncertainty in the estimate. The very wide CI is a result of the study being powered to detect a large effect.

Table 20: Secondary attainment analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention, control)	Hedges' g (95% CI)	P-value
English attainment	345 (45)	10.49 (10.10, 10.88)	347 (24)	10.73 (10.33, 11.13)	692 (345, 347)	-0.020 (-0.14, 0.100)	0.75
Maths attainment	347 (43)	9.35 (8.92, 9.79)	345 (26)	9.61 (9.16, 10.05)	692 (347, 345)	-0.021 (-0.12, 0.077)	0.68

Secondary outcomes on employability skills and attitudes

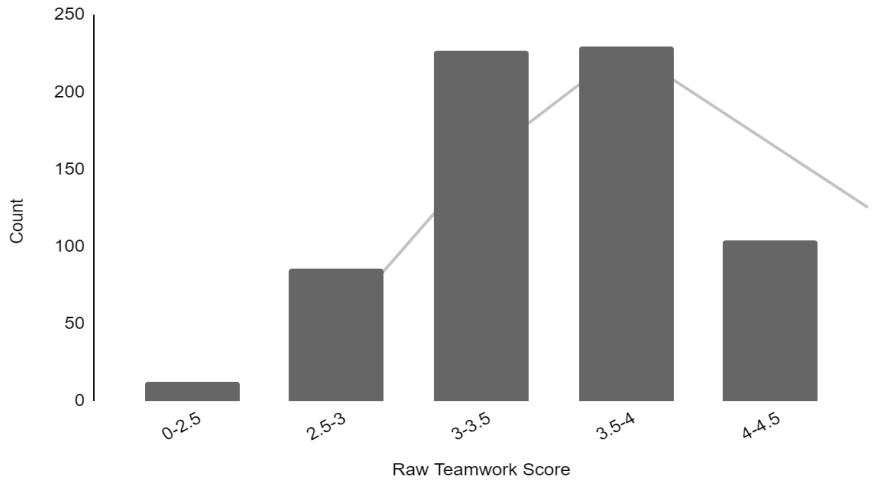
In this section, we present the results for the outcomes collected via our survey that relate to employability skills and attitudes (see Secondary analysis section above for details of the analysis specification).

Teamwork

Teamwork was measured using a 5-point scale, whereby 5 indicates high teamworking skills and 1 represents low teamworking skills, so those estimated effects can take values between -4 and 4. Effects are also presented as Hedges' g to make it easier to compare between outcomes and with other studies.

Figure 6 shows the distribution of outcomes.[36] Scores are distributed across the possible range with a slight skew to the right; we do not have any concerns about floor or ceiling effects in the analysis.

Figure 6: Histogram of teamwork scores



The unadjusted mean in the intervention group is 3.95 and that in the control group is 3.90 (out of 5) (**Table 21**). After adjusting for covariates in the analysis model, the mean difference between the two groups is 0.055, which translates into a Hedges' g effect size of 0.11 (a positive effect) (**Table 21**). This is our best estimate of the effect but, at the 95% confidence level, the results are compatible with effects that range from almost zero to an even larger positive effect. There is therefore, a large amount of uncertainty in the estimate, but the intervention is unlikely to be having a negative effect on pupils' teamwork.

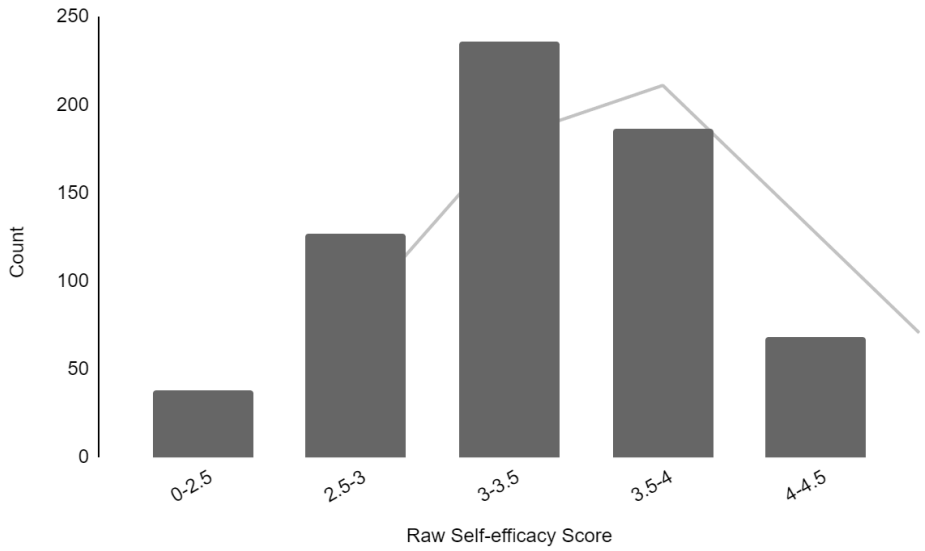
[36] N.B. This graph excludes any counts below ten to be compliant with ONS policies.

Self-efficacy

Self-efficacy was measured using a 5-point scale, whereby 5 indicates high self-efficacy and 1 represents low self-efficacy, so those estimated effects can take values between -4 and 4. Effects are also presented as Hedges' g to make it easier to compare between outcomes and with other studies.

Figure 7 shows the distribution of outcomes.[37] Scores are distributed across the possible range with no observable skew; as such, we do not have any concerns about floor or ceiling effects in the analysis.

Figure 7: Histogram of self-efficacy scores



The unadjusted mean in the intervention group is 3.82 and that in the control group is 3.71 (out of 5) (**Table 21**). After adjusting for covariates in the analysis model, the mean difference between the two groups is 0.11, which translates into a Hedges' g effect size of 0.19 (a substantial positive effect) (**Table 21**). This is our best estimate of the effect but, at the 95% confidence level, the results are compatible with effects that range from very small (but still positive) to an even larger positive effect. There is therefore, some uncertainty in the estimate, but the intervention is very likely to be having a positive effect on pupils' sense of self-efficacy.

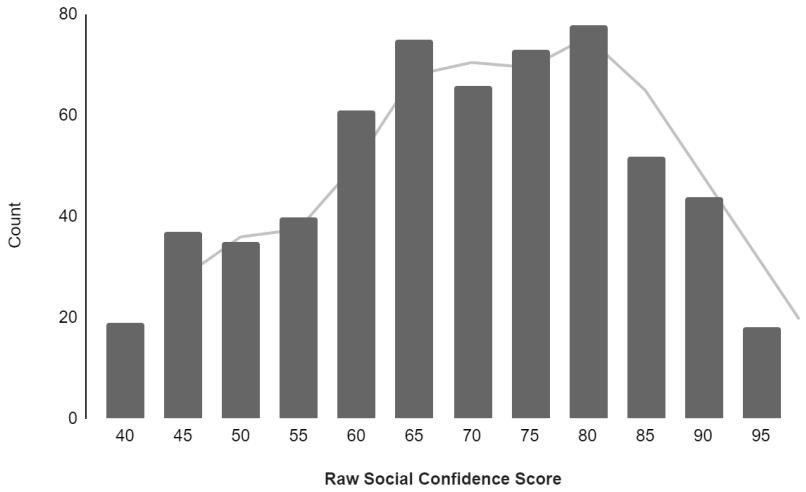
Social confidence

Social confidence was measured using a 100-point scale, whereby 100 indicates high social confidence and 0 low social confidence, so those estimated effects can take values between -100 and 100. Effects are also presented as Hedges' g to make it easier to compare between outcomes and with other studies.

Figure 8 shows the distribution of outcomes.[38] Scores are distributed across the possible range with no observable skew; as such, we do not have any concerns about floor or ceiling effects in the analysis.

[37] N.B. This graph excludes any counts below ten to be compliant with ONS policies.
[38] N.B. This graph excludes any counts below ten to be compliant with ONS policies.

Figure 8: Histogram of social confidence scores



The unadjusted mean in the intervention group is 71.73 and that in the control group is 66.24 (out of 100) (**Table 21**). After adjusting for covariates in the analysis model, the mean difference between the two groups is 8.13, which translates into a Hedges' g effect size of 0.29 (a substantial positive effect) (**Table 21**). This is our best estimate of the effect but, at the 95% confidence level, the results are compatible with effects that range from slightly smaller (but still positive) to an even larger positive effect. There is therefore, some uncertainty in the estimate, but the intervention is very likely to be having a positive effect on pupils' social confidence.

Table 21: Secondary analysis—employability skills and attitudes

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention, control)	Hedges' g (95% CI)	P-value
Teamwork	332 (58)	3.95 (3.90, 4.01)	333 (38)	3.90 (3.84, 3.95)	665 (332, 333)	0.11 (-0.038, 0.26)	0.15
Self-efficacy	311 (79)	3.82 (3.75, 3.88)	312 (59)	3.71 (3.65, 3.77)	623 (311, 312)	0.19 (0.041, 0.34)	0.013
Social confidence	333 (57)	71.73 (69.73, 73.73)	328 (43)	66.24 (64.36, 68.12)	661 (333, 328)	0.29 (0.14, 0.43)	0.000

Analysis in the presence of non-compliance

A pupil is considered to be compliant if they have attended at least the following combination of sessions: five weekly sessions; and one business mentoring session.

Based on the data collected by the delivery team, compliance was high with 78% of participants meeting or exceeding the minimum compliance threshold.

To examine the issue of non-compliance, CACE was estimated. This was done to explore whether low attendance by pupils may be diluting the estimated treatment effect. To calculate the CACE, we use the 'ivregress' functionality of Stata.

Compliance data are available for the full primary analysis sample (n = 689). The high correlation between treatment status and compliance status (r = 0.80) suggests that there is unlikely to be a difference in findings between the ITT and CACE analyses, while the large F statistic of the first stage of the instrumental variables model (F(33, 655) = 153.84) implies that there is no weak instrument issue, so the analysis is meaningful.

The estimated difference in Attainment 8 scores produced by the CACE is 0.15 (95% CI: -2.13, 2.44), which is qualitatively similar to that of the ITT analysis (primary analysis 95% CI: -1.76, 2.01). As with the primary analysis, this is our best estimate of the effect, but there is a large amount of uncertainty in the estimate, as seen in the very wide CI.

The fact that there is some non-compliance does suggest, however, that the average positive effects on self-efficacy and social confidence estimated above, could be increased by increasing participants' attendance on the programme.

Table 22: CACE analysis results

Model	Hedges' g (95% CI)	Sample size	First stage F-test	P-value of treatment variable
Compliance analysis	0.0084 (-0.12, 0.13)	689	153.84	0.90

Missing data analysis

In the Statistical Analysis Plan for this trial, we outlined a missing data strategy. We noted that this strategy would be implemented if more than 5% of data for a covariate in the primary analysis sample were missing or if more than 10% of data for a covariate in a single school were missing.

Investigating the missingness of covariates, we find that FSM and School ID are never missing but Key Stage 2 attainment data is missing for 7% of our sample. So, missing data analysis was conducted for the Key Stage 2 covariates. To assess whether the absence of Key Stage 2 attainment data follows a random pattern (MAR), we modelled the missingness of each variable (Key Stage 2 Maths and Key Stage 2 Reading) using two logistic regressions as specified on page 23 of this report.

For both Maths and Reading, there were no predictors of missingness that were significant at the 5% level (p-values ranging from 0.056 to 0.987). As such, in line with the Statistical Analysis Plan, multiple imputation was not conducted and instead sensitivity analyses were carried out by excluding all covariates from our regression specification for the primary outcome analysis. The coefficient in the sensitivity analysis is directionally different from the primary analysis but the result is substantively the same: the effect is very small, and the 95% confidence interval is very wide (95% CI: -3.34, 2.06).

In the results of the models for missingness of the two Key Stage 2 variables, the p-values for the treatment assignment indicator were 0.706 and 0.382 (for Maths and Reading, respectively), so treatment assignment does not predict missingness in either case. This, in combination with the fact that at the analysis stage we were balanced on all observables, suggests that any unobservable characteristic that leads to missingness is likely to be distributed evenly between treatment and control and hence, not cause a bias in the treatment estimate.

Table 23: Sensitivity analysis results

	Primary analysis				Sensitivity analysis (excluding all covariates)			
Outcome	Regression coefficient	Standard error	Sample size	P-value	Regression coefficient	Standard error	Sample size	P-value
Attainment 8 score	0.12	0.96	689	0.90	-0.64	1.38	739	0.64

No further analysis was carried out for missing outcome data. The extent of missing outcome data and reasons for this missingness are summarised in the 'Attrition' section, above.

Subgroup analyses

Subgroup analyses were conducted to establish whether treatment effects differed for pupils who are, and are not, eligible for FSM. In line with the EEF guidelines, subgroup analysis was conducted both using an interaction effect and a model where the sample is limited to the subgroup of those who did or did not receive FSM.

As **Table 24** indicates, the coefficient on the interaction term (interpreted as the difference in effects between FSM and non-FSM pupils) is small (see Hedges' g for the interaction effect model—equivalent to one month's additional progress). This is our best estimate of the difference in effects between the two groups, however, as with the primary analysis on attainment, the CI on this estimate is very wide (ranging from one month's less progress to three months' additional progress).

When estimating the effects on the FSM and non-FSM subgroups by restricting the sample, those who were not eligible for FSM saw an impact of two months' additional progress (95% CI: [less one month, additional four months]), while for those who were eligible, saw an impact of two months' less progress (95% CI: [less four months, additional two months]).

The point estimates from these two types of subgroup analysis suggest contradictory results. The interaction effect approach suggests that FSM pupils received a slightly larger effect, whereas the restricted sample estimates suggest that FSM pupils received a substantially smaller effect. However, the level of uncertainty around the estimates makes the results inconclusive (and is the likely explanation of this contradiction). We also note in the power calculations above that the study was almost certainly underpowered to detect these heterogeneous effects.

Table 24: Subgroup analysis results

Model	Hedges' g (95% CI)	P-value	N (intervention, control)
Interaction effect	-0.170 (-0.38, 0.043)	0.12	689 (344, 345)
Subgroup (non-FSM)	0.086 (-0.056, 0.23)	0.23	368 (184, 184)
Subgroup (FSM)	-0.091 (-0.26, 0.078)	0.29	321 (160, 161)

Additional analyses and robustness checks

Robustness checks

Key Stage 2 score re-coding

In the primary analysis Key Stage 2 marks were included as continuous variables. As a sensitivity check, we re-code the Key Stage 2 marks as categories. This was not pre-specified in the Statistical Analysis Plan, so this is an exploratory analysis. We do this because both methods have their pros and cons, and we want to confirm that using a different method does not affect the results. We conduct this analysis for all outcomes that use Key Stage 2 marks as a covariate. We find that the substance of the results from the primary analysis do not change as a result of this re-coding.

Table 25: Key Stage 2 re-coding robustness check results

	Pre-specified analysis (Key Stage 2 as continuous)				Robustness check (Key Stage 2 as categories)			
Outcome	Regression coefficient	Standard error	Sample size	P-value	Regression coefficient	Standard error	Sample size	P-value
Attainment 8 score *Scores range from 0 to 90*	0.12	0.96	689	0.90	0.31	0.97	686	0.75
English attainment *Scores range from 0 to 18*	-0.073	0.23	692	0.75	-0.088	0.23	692	0.70
Maths attainment	-0.085	0.21	692	0.68	-0.074	0.20	689	0.71

Scores range from 0 to 18								
Self-efficacy *Scores range from 1 to 5*	0.11	0.043	623	0.013	0.10	0.044	620	0.021

Survey scores re-coding

For all outcomes collected via our survey, it was specified that the analysis would replace any unanswered questions as missing. A sensitivity check was run where outcomes were recoded to only include the average of the questions they answered. For example, if a measure consisted of five sub-questions and someone only answered four sub-questions, the main analysis coded the whole measure as missing for that individual. However, in the sensitivity check, the average over only the four sub-questions the participant answered was taken. It was found that that the pre-specified results are robust to these checks.

Note this was not pre-specified in the Statistical Analysis Plan, so this is purely exploratory.

Table 26: Survey scores re-coding robustness check results

	Pre-specified analysis (unanswered = missing)				Robustness check (only including those they answered)			
Outcome	Regression coefficient	Standard error	Sample size	P-value	Regression coefficient	Standard error	Sample size	P-value
Teamwork *Scores range from 1 to 5*	0.055	0.038	665	0.15	0.049	0.038	679	0.19
Self-efficacy *Scores range from 1 to 5*	0.11	0.043	623	0.013	0.11	0.043	632	0.012
Social confidence *Scores range from 0 to 100*	8.13	2.21	666	0.000	5.60	1.33	679	0.000

Additional outcome analysis

Social confidence subscales analysis

As specified in the Statistical Analysis Plan, since we find a significant difference in the overall social confidence scores across the treatment and control groups, we run exploratory analysis on the confidence subscales. All confidence scores were measured using a 100-point scale, whereby 100 indicates high confidence and 0 low confidence, so those estimated effects can take values between -100 and 100. Effects are also presented as Hedges' g to make it easier to compare between outcomes and with other studies.

As we can see in **Table 27**, the average confidence score in the intervention group is higher than in the control group, for all subscales. The Hedges' g for the scales range from 0.054 to 0.39 and the p-values range from 0.000 to 0.47. The two largest effects are on participants' confidence in speaking to strangers and in public speaking. This makes sense given the programme's emphasis on meeting and working with new adults (business mentors), and on delivering a 'pitch' presentation at the 'Boardroom final'. Conversely, the overall positive effect on social confidence does not seem to be driven by an increase in confidence communicating with friends. The group-based nature of the programme, where pupils are brought together in meetings with peers who they know, but are not necessarily their friends, does seem to be leading to an effect on pupils' social confidence in these respects.

Table 27: Pre-specified analysis of social confidence subscales

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention, control)	Hedges' g (95% CI)	P-value
Public confidence	338 (52)	70.53 (68.28, 72.78)	331 (40)	62.66 (60.34, 64.97)	669 (338, 331)	0.35 (0.20, 0.50)	0.000
Meeting confidence	338 (52)	66.20 (63.84, 68.55)	332 (39)	60.72 (58.52, 62.93)	670 (332, 338)	0.25 (0.10, 0.40)	0.001
Group confidence	340 (50)	75.04 (72.96, 77.12)	331 (40)	70.01 (68.04, 71.97)	671 (331, 340)	0.26 (0.11, 0.40)	0.001
Dyad confidence	338 (52)	75.35 (73.47, 77.24)	331 (40)	71.39 (69.58, 73.20)	669 (331, 338)	0.22 (0.074, 0.37)	0.003
Stranger confidence	339 (51)	60.66 (58.01, 63.31)	332 (39)	50.84 (48.20, 53.48)	671 (332, 339)	0.39 (0.24, 0.54)	0.000
Acquaintance confidence	337 (53)	67.01 (64.53, 69.49)	329 (42)	60.91 (58.51, 63.31)	666 (329, 337)	0.25 (0.11, 0.40)	0.001
Friend confidence	338 (52)	87.69 (86.11, 89.27)	336 (35)	86.87 (85.38, 88.37)	674 (336, 338)	0.054 (-0.093, 0.20)	0.47

Implementation and process evaluation

Fidelity, adaptation, and dosage

The programme structure and resources were closely adhered to by delivery staff and volunteers. All intervention activities: 2 cross-school events; 11 weekly meetings; 4 business mentoring sessions; and *ad hoc* implementation of fundraising campaigns, were delivered for all schools in the trial. In four schools, pupils attended one additional business mentoring session. Pupil attendance figures are presented in the CACE analysis in the section 'Analysis in the presence of non-compliance' above.

While the number and type of sessions delivered were consistent across settings, the approach taken by coaches and business mentors did seem to vary substantially. These variations in facilitation and mentoring style are addressed below under the 'Quality' section.

Quality

Two types of personnel were responsible for delivery of programme sessions: i) the Envision staff member who ran weekly sessions in the school and led the business mentoring sessions out of school (the 'coaches'); and ii) the adult volunteers who supported the delivery of business mentoring sessions in their workplace (the 'business mentors'). Quality of delivery is discussed below in relation to these two roles.

Coaches

The endpoint survey results suggest that the majority of pupils in the intervention group (78%) considered their relationship with their coach to be positive. Around 45% of pupils in the intervention group strongly agreed with the statement: 'I worked well with my Envision coach'. Some substantial variation was seen in quality in the case study sessions that were observed, however. Through interviews and observations, three factors were identified as important determinants of the quality of coaches' delivery: i) the balance of directive and non-directive approaches; ii) support for inclusion; and iii) the nature of the coach–pupil relationship.

Directive/non-directive balance

The coach's ability to strike the right balance between a directive and non-directive approach to session facilitation seemed to be an important factor affecting quality, particularly as the intervention was informal in style and aimed to give pupils substantial ownership over the goal and the means (implying the need for an approach that was not didactic). Pupils appreciated this quality in their coach when they saw it.

> *[The coach] doesn't control everything, [they] just give us ideas and let us decide what to do. So [they're] not like 'You do this and you do this', it's like 'Here's an idea and can you expand on it.'...I prefer that a lot more than just telling us what to do. It gives us more freedom to do what we want.*
> (Pupil 4)

Three key skills were observed as important in supporting coaches to strike this balance. First, while the aim was to give pupils a substantial degree of ownership over their projects, observations of the intervention revealed the importance of coaches taking the right kind of leadership over the process; allowing pupils to have substantial control over their work, but providing a clear structure in which to do this. In the case study setting where engagement and progress seemed to be highest, the coach divided the session into five activities that were long enough to ensure that meaningful work could be done, but short enough to maintain engagement. During each activity, pupils were first left to discuss ideas among themselves, and the coach then circulated to ensure that pupils understood and remained on task. The sequence of these activities was also well-considered, so that pupils made progress towards the overall aim of the session (in this case, to develop their pitch for the boardroom final).

In a setting that appeared to be less successful in terms of pupil engagement and progress, the session was less-well planned. Here, the session seemed to have two different aims: to refine the pupils' pitch ready for the boardroom final, and to support pupils to reflect on their development as a result of participation in the intervention. To address these aims, eight distinct activities were led by the coach. Many of these activities were rushed to ensure that everything could be fitted in, they sometimes seemed unconnected to each other, and pupil engagement decreased as the session

progressed. A lack of preparation from the business mentors in this session also resulted in some lower-quality activities. For example, in one activity business mentors were asked in turn to describe themselves at 15 years old. Their responses to this question were made up on the spot so lacked focus, and the outcome for the pupils was not obvious. The design of some of the activities also meant that only a small number of pupils were actively participating for substantial stretches of the session. During these periods, other pupils started doodling or talking amongst themselves and the coach then began to reprimand them, resulting in a negative, low-energy atmosphere. Careful planning and a well-executed session structure therefore seemed to be a very important factor influencing the quality of experience for pupils. This structured support was also described by pupils as important in enabling them to complete their project tasks.

> *"[Our coach] kept on reminding us about time and making sure that everything was sorted."* (Pupil 14)

The second skill that seemed to facilitate an effective directive/non-directive balance was good questioning. In a session where good progress was observed during boardroom pitch practice, the coach asked questions that encouraged independent thought, often starting with open questions and then gradually probing in a way that helped pupils to identify points for improvement for themselves. For example, one sequence of questions proceeded as follows: 'How do you think that went?; What does that look like?; How will that help in the boardroom?; What do you mean by confidence?; What's the opposite of speaking clearly?; How about the way you came into the room?'

The third important skill observed, which was strongly related to the good use of questioning, was the coach's approach to giving feedback. While pupils described not being led by their coaches, they also highlighted in interviews the value of the feedback given to them by their coaches. This included a mixture of developmental advice and positive reinforcement. In one observation, the coach was particularly effective at facilitating peer-to-peer feedback, providing pupils with useful structure where they were asked to identify a strength in a peer and then suggest how their performance could be 'even better if…' Pupils were familiar with this structure from previous sessions and were able to provide useful and positive developmental feedback to each other as a result. An absence of this good facilitation in another setting resulted in less helpful feedback, from both pupils and business mentors.

Supporting inclusion

Some pupils needed additional support to engage in the intervention; especially those pupils with lower confidence and particularly in the early stages of the programme, when pupils were less familiar with their teammates and business mentors. Coaches were crucial in supporting this inclusion. In successful cases, they used one or more of three strategies. First, activities were structured in ways that made it easier for less confident pupils to engage. For example, in one observed session, the coach used 'snowballing' to structure a brainstorming session. In this activity pupils were first asked to discuss ideas in pairs, then fours, and so on, increasing the size of the group each time. In another activity, pupils were given the chance to write their ideas on a flipchart so that those who were less inclined to speak had an alternative way of contributing. Second, when group discussions were run, the more successful ones were managed by the coach using techniques like 'round-robins' (where each pupil is required to make a small contribution in turn) and gently preventing some pupils from dominating discussions and speaking out of turn. Third, some interviewed pupils suggested that their coach supported inclusion by building the confidence of quieter pupils through positive reinforcement. A less successful strategy was observed in one case, where the coach played music during group discussions, to provide some background noise. While the aim of this strategy was to make pupils feel more comfortable in speaking to their mentors, it seemed to have the unintended consequence of distracting pupils, with conversation leaning towards what song should be played next, rather than the topic of the session.

Coach–pupil relationship

The type of relationship that coaches formed with their teams was described by pupils as different to the usual teacher–pupil relationship.

> *"[The coach] is similar in a way [to a teacher]. [They] still have that authority, but it's not kind of like…I can't find the word. You know how sometimes teachers…are strict and stuff. [They are] strict, but in a, 'You are still a human,' kind of way."* (Pupil 11)

In comparison to their teachers, pupils described their coach as someone they look up to, less-controlling, more relaxed, and easy to talk to.

> *"I think I can talk to [them] about pretty much anything because [they're] not like a teacher. We don't call [them] 'Sir/Miss', we call [them] [coach's name]. [They're] more like a friend, [they] help a lot."* (Pupil 15)

Pupils who had a good relationship with their coach also emphasised the generally positive attitude that they brought into each session. These characteristics combined in the best cases to create a relationship that pupils enjoyed and derived motivation and encouragement from.

Business mentors

The endpoint survey results also suggest that the majority of pupils in the intervention group (75%) considered their relationship with their business mentors to be positive. Around 37% of pupils in the intervention group strongly agreed with the statement: 'I worked well with my business mentors'. In the case study interviews, pupils identified three main contributions from their business mentors: giving practical and developmental advice; supporting fundraising; and supporting them at the boardroom final'. Two factors were identified as affecting the quality of this support from business mentors: i) their ability to give good feedback; and ii) the nature of the mentor–pupil relationship.

Giving feedback

One of the main functions for business mentors was to provide feedback to pupils on their project ideas and on their pitches for the boardroom final'. Pupil interviews suggested that this feedback took the form of developmental advice, positive reinforcement, and practical advice from experience. There was evidence that all three types of feedback were implemented effectively by mentors, with pupils describing positive examples in each category. Developmental advice was given by mentors in particular on the teams' pitches for the boardroom final', and pupils were able to recall points that they were working on in relation to their individual performance and that of their team. In terms of positive reinforcement, pupils described mentors highlighting their strengths after practice pitches, and one pupil interviewed suggested that this had had a positive effect on her confidence in delivering a mock GCSE presentation, outside of the programme. Examples of practical advice cited by pupils include how to work effectively with other people on a project and what fundraising activities are likely to be most effective.

Observations of mentoring sessions suggested that there was room for improvement in the quality of the feedback delivered, however. For example, in one session, the feedback given by mentors after a practice pitch sometimes lacked detail and structure, making it harder for pupils to follow and act upon. In this session, it was not clear that the coach or mentors had a set of clear points that they wanted to convey to the pupils, and they would likely have benefited from using successful structures observed in other cases, such as the 'even better if…' model described above. One mentor suggested during their interview that they were disappointed with the limited amount of time that they were able to give to session preparation due to their work and personal commitments.

Mentor–pupil relationship

As with coaches, the nature of the relationship between mentors and pupils was very different to a teacher–pupil relationship. Pupils described some characteristics of this relationship in similar terms to those described above for coaches; mentors sometimes made sessions fun, they could be easy to talk to and they offered encouragement, especially for pupils who lacked confidence in presenting in front of an audience. This relationship seemed less important to pupils than that with their coach, however. In interviews, some pupils struggled to recall who their mentors were or details of their mentoring sessions. Others suggested that the intervention would benefit from less business mentoring sessions as they did not add substantially to the sessions that were run in school by their coach.

> *"[The mentors] are nice…to speak to, and they are encouraging and gave good advice, which was helpful. But I don't think we got to know them as well as our coaches."* (Pupil 6)

Programme differentiation

Programme differentiation—how distinct the intervention was from existing practice—was assessed through a combination of quantitative and qualitative data. **Table 28** summarises the pupil survey findings on this question.

Table 28: Programme differentiation survey

Item	Control group	Intervention group
Percentage of respondents that reported doing extracurricular activities	60	63
Percentage of respondents that reported doing extracurricular activities that could cause a positive effect on the primary outcome*	58	59
Percentage of respondents that reported doing extracurricular activities that could cause a positive effect on the secondary outcome†	58	59
Percentage of respondents that reported doing youth social action (other than the intervention) as part of their extracurricular activities‡	22	26

* *Activities listed in the EEF Toolkit[39] as having a positive effect on academic attainment.*

† *Activities with a plausible theory of change that link them to one or more of the secondary outcomes (as judged by the researcher).*

‡ *Activities that support young people to take 'practical action in the service of others' (Pye and Michelmore, 2017).*

The survey results suggest that the majority of participants in both the intervention and control groups participated in extracurricular activities, other than the intervention. The results also show that most of these 'additional' activities were of the kind that could have had positive effects on both the primary and secondary outcomes. While there are small differences in the proportion of pupils participating in these activities in the control and intervention groups, these differences are not statistically significant at the 95% confidence level.

The proportion of pupils participating in youth social action activities, other than the intervention, was four percentage points greater in the treatment group than the control group. This is an 18% increase, which seems large, but the difference was not statistically significant at the 95% confidence level.

Therefore, there was no significant difference between the intervention and control groups in terms of their participation in other extracurricular activities. This similarity in conditions between the groups supports the conclusion that the effects estimated in the impact analysis are due to the intervention (and are not an artefact of imbalances in other activities between the two groups).

In terms of their relation to other activities offered to pupils, the intervention's weekly sessions were positioned differently in each of the three case study settings. In setting 1, the sessions were added as an option for extracurricular learning during a weekly one-hour enrichment slot during the school day; in setting 2, they replaced two 30-minute lessons during curriculum time; and in setting 3, they were scheduled once per week for one hour in an after-school slot. In settings 1 and 2, the intervention therefore, formally replaced other activities, whereas in setting 3 an additional slot was scheduled specially for the intervention. The effects from the impact analysis are therefore, likely due to a combination of substitution and additional activity. This analysis was not conducted across all settings however, so it is not possible to quantify the extent to which the observed average effects are due to substitution or addition. It should also be noted that the example of setting 3 is not a simple example of additionality as, even though a new regular time slot was created for the intervention, participation in sessions during this slot may have prevented pupils from participating in other after-school activities; either *ad hoc* ones organised by the school or those organised by other groups.

In interviews, teachers and senior managers identified six types of activity that they considered to be particularly similar to the intervention: career-based (e.g. work experience and interview practice); enterprise (e.g. Young Enterprise club); skills-based enrichment (e.g. those focused on communication skills); individual service (such as litter picking in the school); group service (e.g. food collections for the homeless organised across a year group); and moral education (e.g. a programme of lessons on moral virtues). Despite this broad set of activities that were perceived to be similar, six characteristics were identified as distinguishing features of the intervention. First, the intervention supported pupils to engage with communities beyond the school, through their charity and business partners. Second, the intervention was perceived by some pupils as more interesting than the curriculum-based alternative.

> *"I don't like [our lessons on] virtues because they're just boring, but we go to...Envision...and we talk about our charity."* (Pupil 2)

[39] https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit.

Third, a substantial amount of staff resources (paid and voluntary) are offered by the intervention, providing different types of adult relationships for the pupils and reducing the potential burden on teaching and management staff at the host school.

> *"There is a whole package that you get…[including] the staff that come into school...They have a more informal relationship with the students, and then you have got the corporate mentors, who have a different and more professional relationship. There are other things that you can do [instead of the intervention], you can cope with social action projects, and you can apply for funding for them. But there isn't any support system to make that happen, which obviously Envision provides.*" (Senior manager 3)

Fourth, for some pupils, of all the projects in which they could participate, the Community Apprentice offered them the most ownership. This factor is described in more detail in the section on responsiveness, below. Fifth, the fact that the intervention required participants to plan and deliver a real project distinguishes it from activities that aimed to develop similar skills through simulations such as mock interviews.

> *"The other opportunities involve team-building work and involve working with other people. They involve reflecting on your own development, and those are the similarities. [But the Community Apprentice requires pupils to] plan and organise an event...So, it's not just turning up and doing something, but it's putting something together, delivering it and evaluating it...It's not just personal development. [The pupils have] done something for someone else."* (Senior manager 1)

Sixth, the intervention allows pupils to explore and address a social issue that is important to them, distinguishing it from similar programmes of project-based learning that, for example, support pupils to make money.

> *"[It] is the social impact bit of Envision...that is the different bit about this program[me]. That definitely makes a difference from Young Enterprise and...makes it resonate with us. It makes it transformative for some of the students, particularly because for us, we are a diverse school and it has allowed students to deal with issues that are happening in their communities, in the national news, internationally, in a way that something like Young Enterprise doesn't do."* (Senior manager 3)

The same manager also suggested that this final differentiating feature of Envision's programmes once made them *'trailblazing'* (Senior manager 3), but that there were now more youth social action programmes for schools to choose from. This idea is also reflected in the fact that approximately a quarter of trial participants were taking part in youth social action opportunities other than the intervention.

Reach

Describing reach

The intervention aimed to reach 13 pupils in Year 10 in each participating school, with at least 30% of participants being eligible for FSM in the past six years. These aims were achieved, with all participating pupils coming from Year 10, and 41% being eligible for FSM in the past six years. The selection process for pupils involved two stages. In stage 1, 26 pupils were identified by each school to participate in the trial. This stage was either done solely by teaching and management staff, or it included a process where pupils put themselves forward for the programme prior to the final selection by staff. In stage 2, 13 pupils were randomly assigned by the researchers to receive the intervention.

Effects of the selection process

Three effects of the selection process being only partly voluntary were identified in interviews with pupils and staff. First, both pupils and teachers suggested that a purely voluntary process would have led to groups being formed more on the basis of existing friendships. While pupils did highlight the importance of positive relationships with their teammates, both teachers and pupils also suggested that the random assignment of pupils had the positive effect of forcing pupils to work with new peers. In one example, a pupil specifically signed up to the sample of 26 to be with a friend but was then assigned to the intervention group while his friend was assigned to the control group. Even this pupil reported enjoying the programme and specifically highlighted the opportunity to work with new people as beneficial. Team relationships emerged as an important theme across the IPE, with further implications discussed below with regards to the mediating and moderating factors of the intervention, and participants' responsiveness to the intervention.

Second, the fact that teachers had control over the first stage of selection meant that they could select pupils who they thought would benefit from participation but would not necessarily be motivated to sign up in a purely voluntary process. This seemed to result in some pupils participating who were not motivated to do so other than their teacher's encouragement.

Third, one senior manager thought that buy-in was low among some pupils because they felt that they had been told to participate by the teacher, rather than choosing to do so. This effect was thought to be mitigated by including an application process for pupils in stage 1, but seemed to be a particular issue in the case study setting that chose to run the weekly sessions in an after-school slot (which requires higher pupil motivation). Conversely, in a different setting, the lead teacher thought that attendance at intervention sessions was higher than it would be, because a purely voluntary selection process would give pupils the impression that they need not attend.

Pupil motivations for participating

The rate of participation in the programme was high with 78% of participants meeting or exceeding the minimum compliance threshold. Pupils had varying types of control over their participation (due to the range of approaches taken to pupil selection), but to the extent that participants chose to participate, three categories of motivating reasons were identified: i) social reasons; ii) altruistic reasons; and iii) reasons relating to personal benefits. In the first category were participants who were motivated due to the influence of others. For some, the intervention offered the chance to spend time with friends, and for others it was a chance to interact with new people. The positive reputation of Envision among previous participants was also a motivating factor for some. In the second category were participants who signed up to help others. Members of this category described wanting to support charitable work and to make concrete positive changes to peoples' lives. This provides support for the idea, expressed in the intervention's logic model, that self-efficacy is an important motivating factor for behaviour. The third category captures a set of benefits that pupils thought that they could accrue for themselves through participation: the opportunity to develop self-confidence, something to put on their CV, the potential to win prizes, and the opportunity to miss lessons and go on trips.

Factors affecting participation

From interviews with staff and pupils, six factors were identified as potential barriers/enablers to participation. First, the intervention coincided with mock GCSE exams in some settings and with other extracurricular responsibilities (e.g. the school play) for some pupils. Both pupils and teachers identified this as a barrier to participation, with the level of impact on the pupil being partly dependent on their ability to manage competing priorities.

> *"[I]t was during exam time and it is stressful, and I have to be really organised, like over organised and plus it was too much pressure. So, maybe [the programme should run at] a different time."* (Pupil 9)

A second-related factor was that core intervention activities sometimes ran outside of school hours. This was identified as a particular barrier for Muslim pupils (who attend Mosque after school) and pupils who have to travel home further. There were also some pupils who just expressed a general aversion to participating in activities after school.

Third, support from the school was identified by business mentors and school staff as an important enabler. Schools that had sufficient flexibility in their timetable to allow for weekly sessions to run during the school day made it easier for pupils to participate. Schools that were proactively managing extracurricular opportunities across departments, and were systematically monitoring pupil participation, also felt better able to encourage the right pupils to participate in the intervention.

> *"[The important thing is] communication between the pastoral and careers team. We have a tracker this year which is a simple Excel spreadsheet so we can see which opportunities each child has done…So, if there is an opportunity to go on a trip or take part in a workshop we can see who hasn't done anything in those categories, and then think on who we can target."* (Senior manager 1)

One case study school occasionally struggled to offer logistical support, such as accompanying pupils to external events and signing off risk assessments for trips related to the intervention and this created barriers to pupil participation in these cases.

Fourth, support from family members was also cited as important by some pupils. For some, parents or older siblings encouraged them to participate, and one pupil described her parent providing practical support for her project by doing some baking for a fundraiser.

Fifth, the level of self-confidence that pupils felt prior to participation was described as an important factor. For example, one pupil with low self-confidence did not want to take part in the programme at the beginning of the year, even though they could describe clear potential benefits for themselves and had encouragement from their family. In particular, this pupil was worried that the intervention would require them to speak in front of other people. Pupil confidence with communication is a key theme in multiple parts of the IPE analysis; it is a target secondary outcome for the intervention but has also been identified as a potential moderating factor, and an important factor for coaches and mentors to address through quality facilitation of the intervention.

The sixth factor identified as influencing pupil participation was the way in which the intervention was initially described to pupils. Schools that were more successful in this regard, gave clear presentations to the potential participants that described the aims, activities, and potential benefits of participation, sometimes using previous participants to help. In other cases, pupils described finding it difficult to understand what the intervention was about, particularly based on the information provided in the letter home to parents, so were unsure as to whether they wanted to participate.

> *"I didn't originally understand what it was. In the letter it talked about who our business mentors are going to be and ours is [organisation name] and I was a bit confused about what we would be doing with [organisation name] because I don't want to get involved with [that type of organisation]. Then I wasn't sure if it was going to take out time [from] my lessons. I get nervous when I get taken out of lessons because I just like being able to learn the stuff there and don't have to catch up at home or miss out on stuff."* (Pupil 13)

Responsiveness

Pupil engagement

Evidence of high-pupil engagement was seen during some observations. For example, during an in-school session, led by an Envision coach, pupils demonstrated detailed recall of activities, of developmental feedback they had received from their business mentors, and of their peers' personal development goals; in group discussions, pupils responded to questions from the Envision coach while others listened; and the team pitches for the boardroom final included contributions from all team members. Engagement was lower during some observed sessions, and amongst some pupils, however. In one session, some pupils seemed anxious and found it difficult to engage in the predominantly discussion-based activities. In a business mentoring session that was observed, pupil attention waned as the session progressed, with some pupils seeming to switch off entirely when business mentors were delivering feedback to the group, doodling and talking amongst themselves. This range of pupil engagement also showed up in interviews. The most engaged pupils were able to describe their experiences of the intervention in detail and could give examples of a substantial amount of additional work that they had done on their project outside of the core sessions. At the other end of the spectrum were pupils who had limited memory of the year and had made much smaller contributions to their teams' fundraising projects.

Factors affecting pupil engagement

Four factors were identified from observations and interviews as affecting pupil engagement: i) the nature of the activities; ii) pupil ownership; iii) pupils' relationships with their teammates; and iv) winning or losing the boardroom final. These four factors are discussed in turn below. (Observations and interviews suggested that the quality of intervention delivery also varied substantially, affecting pupil engagement in some cases, but this is discussed in the section on 'Quality' above.)

Nature of the activities

Four features of the intervention's activities seemed to affect pupil engagement. First, the informal style of sessions seemed to have both positive and negative effects on pupil engagement, depending on how it was implemented. Interviewed pupils reported enjoying the relaxed atmosphere in sessions, the freedom that they were given, the emphasis on creativity, the fact that not much reading and writing was required, and the use of games and pizza. These features distinguished the Community Apprentice from normal lessons, especially subjects like Maths. When this

informality was well-managed by the Envision coach, it was combined with well-structured sessions and high-quality facilitation, having a positive effect on engagement. Pupils also spoke about the importance of progression during the programme, from more 'fun' beginnings to more serious and purposeful later sessions.

> *"In the beginning it was just getting to know each other, getting to know what it's about and playing some games. The first session was just relaxed and then we started choosing our charities and we had votes and it gradually became more about that, more focused on raising money for the charity."* (Pupil 6)

When sessions were less well-planned and facilitated, this was sometimes associated with pupils disengaging, some low-level disruption, and less progress being made in sessions. Pupils, teachers, mentors, and senior managers all suggested that boys found the informality of the sessions particularly difficult to deal with, sometimes disrupting sessions and sometimes not completing their assigned tasks as readily as the girls, who were described as more mature and therefore, more capable of engaging in this type of intervention. In one setting, the lead teacher and senior manager identified the skills of managing the right mix of informality and structure as key to the success of the intervention. Another lead teacher described how the school's approach to behaviour management overall made it difficult for some pupils to engage in the Community Apprentice effectively.

> *"[W]e are very structured with the students in terms of lessons, line ups, the day for a student is very formal. I think some of the students found it difficult in how informal Envision was."* (Teacher 1)

Second, the fact that some sessions took place after school was a barrier to engagement for some pupils, either because they would rather be doing something else with their free time, or because they were tired from the school day so found it harder to concentrate.

Third, some students reported disengaging at points where they thought the programme was repetitive. Some identified repetition in the weekly in-school sessions; for example, presentation practice, which took place over a number of weeks. Others described having to write the same plans down twice; once in a school session and once with their business mentors, which they were not inclined to do.

Fourth, 'mini challenges' and 'learning points' were built into the programme curriculum, linked to a series of badges that pupils were awarded for completing a set of tasks or demonstrating a particular skill. These learning points contributed to the team's score for the inter-school competition. These seemed to have a positive effect on pupil engagement, acting as an incentive for pupils both in themselves and, for some, as a way of competing with peers in their team.

Pupil ownership

As part of the programme, each team of pupils was asked to select a local charity to support with their fundraising. This ownership over the goal of the project was identified by both pupils and senior managers as important for pupil engagement because it allowed them to work on a cause that they felt personally attached to or considered to be important.

> *"It was quite a personal charity for us because...a lot of us walk through town to get to school, and there are lots of homeless people around so it affects us quite personally, and the money that they [the homeless charity] use it was going to be used usefully and we just thought it was a good idea."* (Pupil 1)

Some pupils described the event at which they picked their charity as a turning point, with sessions feeling more purposeful and enjoyable afterwards. For some, this sense of purpose and increased motivation was further enhanced by the setting of a fundraising target. Ownership over the means was also important for some. Pupils described two types of 'means ownership': deciding what type of things they would do to raise funds; and having some control over team meetings. For example, in one case study setting, pupils were selected to be 'group leaders' each week, with some responsibility for facilitating the discussions and recording ideas on the board. Both goal and means ownership were enhanced by simple democratic decision-making processes.

> *"[M]any of us had different ideas…Some of us wanted to do a different charity, but we did vote on it and we discussed it out and we did debate a bit. We did come to a conclusion and everyone was happy."* (Pupil 15)

Pupils' relationships with their teammates

The intervention required pupils to work closely with peers in every session, with the majority of session activities involving paired and group discussions and the fundraising projects themselves requiring pupils to work together to complete tasks. Pupils described the dynamics and relationships in their teams as important for their engagement, the effectiveness of their projects and their development as individuals. Three characteristics of a good team were identified by pupils as important. First, pupils described diversity as being valuable in terms of levels of confidence, types of interests, and competency-based strengths. Pupils suggested that had they been left to choose their teams, rather than being randomly assigned, they would have defaulted to friendship groups, which may not have been so effective.

> *"It was really fun. I made a lot of new friends and I'm glad that I got put into that group. Because if I got put into a group that I picked, we probably would have had some clashes in personality and things like that. And because it was loads of different people we just came together really well*." (Pupil 11)

Second, the fact that groups were diverse in these ways and required so much verbal interaction, meant that pupils with lower confidence found it difficult to engage at times. This was particularly the case near the beginning of the programme, when quieter pupils described feelings of anxiety about their participation. Both these pupils and the more confident ones highlighted the importance of creating a feeling of safety for less confident pupils to engage. With this aim in mind, pupils described learning to be positive with each other (giving praise, for example) and making a conscious effort to give everyone the chance to speak during group discussions. The quality of support and facilitation from the Envision coach was also important in this regard, as discussed in more detail below. Third, the assignment of clear roles and responsibilities also supported pupil engagement.

Winning or losing the 'Boardroom final'

The competition element of the intervention was clearly a motivator for some pupils. Others, however, explicitly said that winning the competition was not important to them, and that they were driven by the cause that they had selected and/or the fundraising target that they had set themselves. These latter pupils did however recognise the benefit of the competition as a motivator for individuals who were less concerned with the cause.

Winning and losing did seem to have strong effects on some participants, with pupils from one of the winning teams describing feelings of euphoria when the results were announced at the 'Boardroom final'.

> *"At the end we were all screaming!"* (Pupil 2)

In addition to this, these pupils described receiving extra esteem from their teachers when returning to school—with recognition of the victory being posted on the school website, for example—and pride from their families. Conversely, pupils on losing teams were genuinely upset, either because they were particularly invested in the competition part of the programme or because they felt that they had let down their charity partner. One pupil also described feeling intimidated by the competition.

> *"I felt like it was a bit competitive, and…to be honest felt a bit intimidated because everyone seemed much older than us."* (Pupil 12)

These negative effects of losing were mitigated in some cases by feelings of satisfaction that came from raising money for charity, making new friends, participating in enjoyable activities, or from winning alternative awards as part of the programme. The extra awards were given out to individuals and teams who were judged to have shown the most development in line with the intervention's target outcomes.

School engagement

While school staff were not responsible for running any of the programme sessions, senior management needed to sign-off on their school's participation, and the lead teacher was required to support the day-to-day logistics of the intervention. Four factors seemed to influence the engagement of school staff: i) the level of alignment between the intervention's aims and the school's ethos; ii) the sign-up process; iii) the perceived costs and benefits of participation; and iv) the timing of activities.

Alignment with school's ethos

A strong interest in character education and supporting active citizenship at senior management level seemed to contribute to an engaged and supportive attitude from school staff.

> "There is a real synergy between what Envision stands for...and what we want to develop in the young people in our school. We are passionate about encouraging young people to care about the communities that they live in, and to not just be individuals and wanting to achieve things for themselves but seeing themselves as part of something that is much bigger. We think Envision encourages young people to do that well. We also think that genuinely, young people can make a difference, and if you take the time to help them develop the skills to be able to do that, they can change the world. I think Envision gives them an opportunity to do that." (Senior manager 3)

In one case study school, where such overt enthusiasm for the aims of the intervention was absent, and where the lead teacher described a stark difference in the pedagogical style (see section on 'Pupil engagement' above), the school also seemed to have the most issues in supporting implementation.

Sign-up process

The nature of the sign-up process for schools seemed to be an important determinant of their subsequent engagement. Two elements of this process emerged as particularly important. First, schools that had worked with Envision previously (through their post-16 programme) described being committed to their general approach and trusting in the quality of their work. Some teachers and managers also felt that this familiarity with the post-16 programme made it easier for them to support the intervention. However, for one school, their prior engagement with the post-16 programme seemed to lead to false expectations about the intervention.

> *"We struggled the first time around, to be honest we didn't quite realise how big it was going to be. I think something that we missed on our part was that with [Y]ear 12 [the programme] is very self-sufficient; so the students apply and take part and...they can go on their own because they are old enough. We just thought great, that will run the same way in [Y]ear 10. What we found very difficult was because we offer so many different opportunities, it was just trying to squeeze everything in and there were times where it was difficult."* (Senior manager 1)

When the intervention ran with a second cohort in this school, they made a point of clarifying expectations with Envision; both in terms of the logistical support required form the school and the level of structure that school staff believed their pupils needed. The school also suggested that their misunderstandings on these points were as much down to their own approach to the sign-up process than Envision's; for example, acknowledging that they were sometimes slow to respond to logistical information. Second, one lead teacher highlighted the importance of having the right staff involved in the set-up conversations with Envision. In this school, the member of staff who was asked to take a lead on supporting the programme was absent from the early meetings with Envision. While the intervention ran smoothly in this setting, this lead teacher described having to put in extra effort to get themself up to speed, which they would rather have avoided.

Perceived costs and benefits

Interviews with teachers and senior managers revealed an implicit cost-benefit analysis that was sometimes being undertaken when deciding whether and how much resource to invest in the intervention. Both teachers and managers highlighted the future cost of the intervention (which was free during the trial) as potentially prohibitive. To justify spending money on the programme, staff suggested that they would expect quantitative evidence of its positive effects (particularly in terms of attendance and academic progress) and would probably want more pupils to be able to participate. One senior manager suggested that the marketplace of youth social action providers was now much larger than when Envision first came to the school, and that the school was considering other providers to ensure that they were getting the best value for money.

Some schools were less clear on the benefits that they expected their pupils to accrue from participation, however. In the case study school that was less aligned in terms of ethos and teaching approach, and that engaged less in the sign-up process, staff seemed less concerned with the specific outcomes of the intervention.

> *"I just thought a 10-week programme, students will go along, get something out of it and do something."* (Teacher 2)

Timing of activities

This less engaged school was also keen for the intervention not to interfere with activities during the school day. Running the weekly sessions after school avoided taking students out of lessons that they considered to be more important, and it allowed the lead teacher to offer more support. This reluctance to take pupils out of lessons sometimes made it difficult to attend business mentoring sessions and cross-school events on time or for the full sessions. The other case study schools did not experience these kinds of difficulties however, with one school giving two short lessons a week over to the intervention, seeing it as an acceptable alternative to their 'virtues curriculum', which was the only curriculum that was replaced.

Logic model

The impact analysis has shown that the intervention can have a positive effect on participants' employability skills and attitudes (especially their general self-efficacy and social confidence), but not their academic attainment. Beyond these outcomes that were listed in the pre-evaluation logic model, the IPE has identified a series of unintended outcomes—in the form of attitudes, behaviours, and skills—that some pupils seem to have experienced. All but one of the mediating mechanisms specified in the logic model also seemed to be present in some cases and, beyond these, three unintended mechanisms were also identified, helping to develop our understanding of how the intervention works. The absence of one key mechanism may also help to explain why no effect was seen on the primary outcome. In terms of the factors that potentially moderated the effects of the intervention, four categories emerged as particularly important: i) the quality of the coach; ii) baseline pupil characteristics; iii) team dynamics; and iv) school characteristics.

Unintended outcomes

Interviews with staff and pupils revealed four outcomes—action-orientation, resilience, ability to organise themselves, and participation in future social action—not specified in the logic model, that may have been felt by some pupils. First, senior managers described some pupils as becoming more 'action-oriented' as a result of their participation. The intervention, as they described it, had given pupils a chance to instigate a project, generating ideas for fundraising activities, and to see it through to completion. Pupils described a similar positive effect on their general attitude.

> *"I want to do more things [outside of the intervention]…If there's an opportunity I'll actually do it. Because at first I didn't want to do it, and I did it, and I helped, and I felt it was a good decision so I do more*." (Pupil 2)

Second, related to this increased willingness to participate and take action was, for some pupils, an increased 'resilience'. Pupils described participation as hard work at times, particularly when they had to complete tasks outside of the weekly sessions, and also described setbacks, such as fundraising events that failed to raise any substantial funds. Third, pupils described an increased sense of self-esteem that came from helping people through their project and, for some, this was what they described as the primary benefit of participation.

> *"First of all, it's just like...a good feeling. It makes you feel nice because you are helping the homeless people and you are making money, and…it just makes you feel good about yourself."* (Pupil 5)

In terms of skills, pupils highlighted their increased ability to 'organise themselves' and 'manage their time'. Senior managers and pupils also described an enhanced ability to solve problems both individually and collectively (through voting, for example).

Finally, pupils described one key behavioural outcome that seemed to result from participation: more participation in social action. For some, the intervention had encouraged them to take part in further fundraising activities with other organisations, while others reported considering careers with charities in the future. This is one of the long-term goals of the programme, which is why an attempt was made to conduct secondary analysis on the outcome (see subsection of 'Impact evaluation' on 'Secondary research questions' in section 'Evaluation objectives' above). It is classified here in the findings as an 'unintended outcome' because it was not included in the evaluation logic model, but this was just a decision of presentation.

Mediating mechanisms

The mediating mechanisms in the pre-evaluation logic model that link the intervention's activities to the secondary outcomes all appear above in the description of the factors that seemed to affect pupil engagement. As well as supporting engagement, these mechanisms—pupil ownership, the completion of mini challenges, achieving goals, winning the competition, and forming positive relationships with teammates—all seemed to contribute to the secondary outcomes. Beyond these hypothesised mechanisms, which did seem to play out for some pupils, four other mechanisms were identified in interviews with staff and pupils.

First, the intervention seemed to place pupils in situations of controlled discomfort as a way of improving their social confidence. These situations included small group discussions with new peers at school which, as described above, was difficult for some pupils at first. They also included multiple practice pitches in weekly sessions and in front of the teams' business mentors. Pupils described feeling anxious in these situations, but also satisfied and more confident after they had got through them. The fact that some of the intervention activities took place in environments like local businesses, that were new or unusual to pupils, added to this sense of initial discomfort, and to the eventual pay-off. One missed opportunity that was identified in this regard was at the cross-school events, where inter-school mixing seemed limited, with a lack of support to help participants overcome their reluctance to engage with each other.

Second, and relatedly, vicarious experience seemed to play a role in social confidence development for some pupils. Some pupils described seeing their peers developing their presentation skill and confidence and this having a positive effect on their confidence. Others described both positive and negative effects of watching teams from other schools delivering their pitches at the final event. Where pupils saw another group struggle with nerves, it increased their confidence, where they saw pupils that they considered to be stronger, it could have a negative effect (as for the pupil who described feeling intimidated above).

Third, while actually achieving the project goals (raising money and winning the competition) did seem to be an important mediator of self-efficacy, the perceived difficulty of the task also seemed to be important for some. Fourth, both senior managers and pupils described the importance of the intervention feeling like a real-world experience. The features of the intervention that seemed to produce this feeling were the fact that pupils met and selected their local charity (being pitched to for support), the interaction with professional adults in real workplaces, and the fact that they were addressing issues that were visible to them in their local area (e.g. homelessness).

The mechanism that did not appear in the IPE data was an improvement in participants' attitude towards education. In the pre-evaluation logic model, this was hypothesised to be the bridge between the secondary outcomes and the primary outcome. The fact that no improvement in this outcome was observed by the researchers during case study visits, or described by any of the interviewees, may help to explain why no effect was seen on the primary outcome.

Moderating factors

In terms of the factors that potentially moderated the effects of the intervention, four categories emerged as particularly important: i) the quality of the coach; ii) baseline pupil characteristics; iii) team dynamics; and iv) school characteristics. All of these items have been discussed above in relation to pupil and school engagements with the intervention. Interviews suggested that, beyond what has already been discussed, general motivation at baseline was also an important pupil characteristic. As the intervention relied on pupils to take the initiative to some extent—setting goals and performing tasks for their fundraising campaigns—its success seemed to be dependent on having a critical mass of pupils with the motivation to contribute substantively to the project. Pupils suggested that their teams often relied upon certain individuals to take a lead, particularly in the beginning. Pupils and teachers also described occasions when peers failed to complete a task, putting the project at risk. In the most extreme case, a teacher felt that she had to complete the majority of the baking for a bake sale, because she was concerned that the pupils would not, highlighting the importance of practical support from school staff, particularly when pupil motivation is lower.

Revised model

Based on these findings, and those that preceded in the sections on responsiveness and quality, a revised logic model is presented below (**Figure 9**). In summary, the following changes are proposed to the logic model.

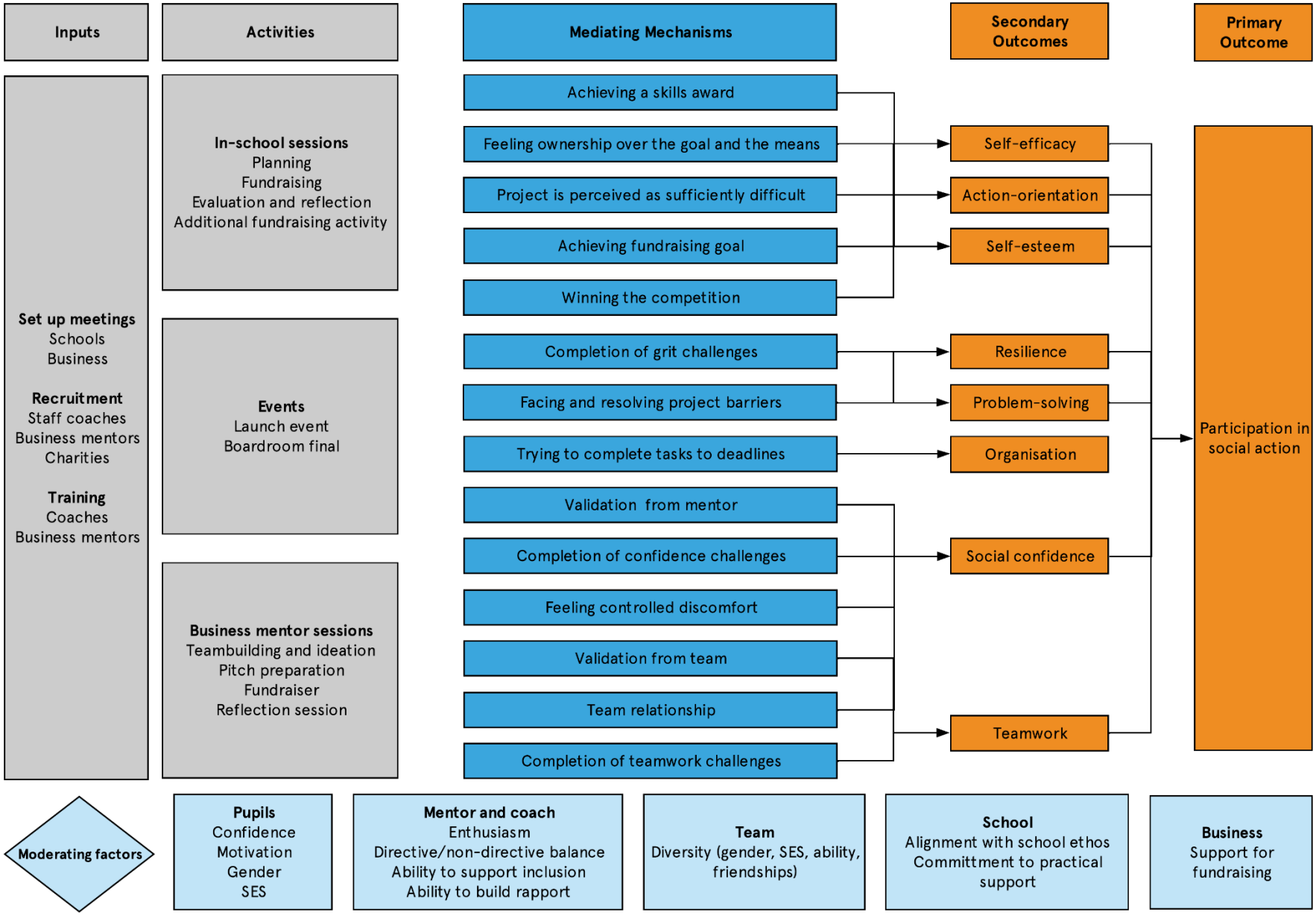
- The unintended outcomes described above have been added. Participation in more social action (the behavioural outcome) has been positioned as the primary outcome to reflect the real primary aim of the

intervention (which was not designed to improve academic attainment).[40] The other unintended outcomes identified have been positioned as additional secondary outcomes rather than mediating mechanisms as they could be seen as ends in themselves. This presentation of outcomes is for completeness. This version of the model presents the best explanation we have of what the intervention aims to achieve, and how it might do so. In any future evaluation work, it would not be advisable to attempt to measure all these outcomes for statistical reasons, but this more complete logic model provides a longer list to choose from, depending on the research question.

- The unintended mediating mechanisms described above have been added.
- The mediating mechanisms have been reorganised and, in some cases, re-described to better fit the category definitions (e.g. the coach's ability has been repositioned as a moderator), and to better align with the details of the IPE findings.
- The moderating factors have been substantially reduced to include only those that appeared in the data.

[40] Envision's aims have since changed and, at the time of writing (August 2024), participation in more social action is no longer the primary aim of the programme, which is now more focused on reducing the number of young people not in education, employment, or training.

Figure 9: Revised logic model



Cost

The financial cost of the intervention in a school is summarised in **Table 29** below. This shows that the majority of the financial costs of delivering the intervention during the evaluation period were incurred by Envision and subsidised by the EEF and a range of other funders (including trusts, foundations, and corporate partners). The total cost per school per year of delivering the intervention is £5,970.76 (equivalent to £459.29 per pupil per year). The intended business model for the intervention in the future includes a fee to schools of £500 per year, with the remaining costs incurred by Envision and paid for by trusts, foundations, and corporate partners. Outside of the evaluation, it is therefore, estimated that the financial cost incurred by schools running the intervention will be £604.10[41] per year (or £46.47 per pupil per year).

Table 29: Cost of delivering the Community Apprentice in a school

Item	Type of cost	Cost per year	Total cost over three years	Total cost per pupil per year over three years
Materials and equipment (e.g. printing)	Running cost per school	£6.50	£19.50	£0.50
Travel and subsistence	Running cost per school	£97.60	£292.79	£7.51
The EEF cost subsidy per school	Running cost per school	£1,066.67	£3,200.00	£82.05
Subsidy from other sources	Running cost per school	£4,800.00	£14,400.00	£369.23
Total		£5,970.76	£17,912.29	£459.29

Table 30 below, shows the estimated cumulative costs for a school running Envision over a three-year period. No one-off start-up costs were identified by schools, so the cost each year is the same and the cumulative costs increase linearly.

Table 30: Cumulative costs of the Community Apprentice (assuming delivery over three years)

	Year one	Year two	Year three
Community Apprentice	£5,970.76	£11,941.53	£17,912.29

Table 31 below, shows the estimated time spent per school, by school staff on the intervention, broken down by activity. There were no additional resources that schools needed as prerequisites for participation that they would not otherwise be expected to have. The intervention can be delivered in normal classrooms or meeting rooms, and no special equipment (such as information and communications technology) is required. Travel was an important part of the intervention, but schools dealt with this using their existing systems; either mini buses that were routinely used for pupil travel or public transport.

Table 31: Time spent by school staff on the Community Apprentice

Activity	Time spent (hours)
Delivery in school day	16
Delivery out of school day	13
Preparation	7
Supply cover	1
Other (e.g. extra meetings)	1
Total	38

[41] This is the cost of the £500 fee plus the cost of materials and equipment (£6.50) and travel and subsistence (£97.60).

Notes on the calculations

Of the 30 schools participating in the trial, 26 completed cost evaluation interviews. For each of the items in the tables above, a mean average was calculated across all respondents. Not all of the 26 schools were able to give an estimate of the cost of each item listed. If a school was unable to provide an estimate, they were counted as a non-respondent and excluded from the mean average calculation for that item. Where respondents gave estimates in terms of days, this has been converted into hours by assuming a seven-hour school day.

Conclusion

Table 32: Key conclusions

Key Conclusions	
1.	Pupils in the Community Apprentice intervention group made no additional progress in their GCSE Attainment 8 score compared to pupils in the control group. This is our best estimate of impact, which has a moderately high security rating. However, there is a high level of uncertainty around the result: the possible impact of the programme could range from a negative effect of two months' less progress to a positive effect of two months' additional progress.
2.	Pupils in the intervention group showed improvements in employability skills and attitudes, particularly self-efficacy and social confidence, compared to pupils in the control group. While there is some uncertainty around effect sizes for these outcomes, findings suggest that the programme is likely to have had a positive impact on self-efficacy and social confidence, in line with the programme's aims.
3.	Subgroup analysis on whether the programme was more or less effective for FSM-eligible pupils is inconclusive due to the small sample size.
4.	The programme was mostly delivered as intended, suggesting that the estimated effects are likely to be due to the programme design. However, the IPE found a high level of variation in the quality of delivery by coaches and mentors. Envision coaches were found to have an important impact on quality, especially their ability to strike a balance between directive and non-directive session facilitation. More consistency in the quality of coaching would likely lead to greater positive effects.
5.	The logic model hypothesised that the programme would have a positive effect on pupils' attitudes towards their education. No evidence was found to support this, which might explain the lack of effect on academic attainment.

Impact evaluation and IPE integration

Evidence to support the logic model

There is strong evidence to support the following components of the logic model:

- The activities: While there was some variation in quality of coaching, the fidelity data collected suggests that the intervention was delivered as intended.
- *The mechanisms:* The majority of the mechanisms found support in the interviews and observations conducted.
- *The secondary outcomes:* The impact analysis suggests substantial positive effects for at least pupils' sense of self-efficacy and their social confidence.

The evidence from this study suggests that the hypothesised moderating factors were less well identified, with a substantially reduced set presented in the revised logic model post-trial. The IPE findings suggest that four main factors were important: i) the quality of the coach; ii) pupil motivation; iii) team dynamics; and iv) school staff willingness to provide pupils with practical support to implement their fundraising projects.

The evidence relating to the primary outcome is inconclusive, with very wide CIs on all impact estimates that relate to academic attainment. The IPE has also identified some potential unintended mechanisms and outcomes. As such, we have proposed a revised logic model that does the following.

- The unintended outcomes—action-orientation, resilience, ability to organise themselves, and participation in future social action—have been added. Participation in more social action (the behavioural outcome) has been positioned as a primary outcome alongside academic attainment as it seems to reflect the real primary aim of the intervention (which was not designed to improve academic attainment). The other unintended outcomes identified have been positioned as additional secondary outcomes rather than mediating mechanisms as they could be seen as ends in themselves. This presentation of outcomes is for completeness. In any future evaluation work, it would not be advisable to attempt to measure all these outcomes for statistical reasons, but this more complete model provides a longer list to choose from, depending on the research question.
- The unintended mediating mechanisms—feeling controlled discomfort, vicarious experience, perceived difficulty of the project, and feeling like a real-world experience—have been added.

- The mediating mechanisms have been reorganised and, in some cases, re-described to better fit the category definitions (e.g. the coach's ability has been repositioned as a moderator), and to better align with the details of the IPE findings.
- The moderating factors have been substantially reduced to include only those that appeared in the data. They relate to characteristics of: i) the pupils; ii) the mentors and coach; iii) the team; iv) the school; and v) the supporting business.

The inconclusive evidence in relation to academic attainment is likely to be due to an insufficient number of pupils in the sample (the study could have been underpowered), as described in the sample size calculations. Future research could address this by re-running the trial with a substantially larger sample. A larger sample is also required to get a good estimate of any differential effects for pupils eligible for FSM (identifying whether the intervention can be a 'gap closer'). Future research could attempt to estimate the effects on the potential unintended outcomes that we have identified.

Interpretation

This study provides strong evidence that the Community Apprentice programme has positive effects on pupils' sense of general self-efficacy and their social confidence. There is also some evidence in support of a positive effect on teamwork.

We cannot say what kind of effect, if any, the programme has on academic performance due to the large amount of uncertainty in the results. However, there are four reasons to suggest that the programme does not improve academic performance. First, our best estimate of the impact from this study is that there is no effect. Second, the intervention was not designed to improve academic performance, and we found no evidence in the IPE to support the idea that the intervention improves pupils' attitudes to education (the key hypothesised mechanism in this regard). Third, we have found no strong evidence in the wider literature to suggest that youth social action programmes like this can have a positive effect on grades. Fourth, the intervention takes pupils out of some academic lessons, which would likely have been beneficial to their academic attainment had they attended them. It therefore, feels unlikely that the programme does improve attainment.

The programme was delivered with high fidelity in terms of adherence by delivery staff to the core components, so we can be confident that the estimated effects are coming from the programme as it has been designed. There seems, however, to be a broad range of delivery quality in terms of coaching. The Envision coaches are particularly important, especially their ability to strike the right balance between a directive and non-directive approach to session facilitation. More consistency in the quality of coaching would likely lead to greater positive effects.

There was no significant difference between the intervention and control groups in terms of their participation in other extracurricular activities. This similarity in conditions between the groups supports the conclusion that the effects estimated in the impact analysis are due to the intervention (and are not an artefact of imbalances in other activities between the two groups).

These results build on the existing literature in two main ways. First, we provide the first robust evidence—in the wider literature and for the EEF Teaching and Learning Toolkit[42]—on the effects of youth social action on academic performance. Second, we add to the growing body of literature that identifies a positive causal effect of youth social action programmes on non-cognitive skills and attitudes (Cameron *et al.,* 2017; Fitzpatrick *et al.,* 2021; Kirkman *et al.,* 2016). Our study suggests that the largest effects are on young people's social confidence and general self-efficacy; a very similar finding to the latest academic work on the topic that combined quasi-experimental impact findings with in-depth qualitative research to build a theory of how youth social action works (Taylor, 2021).

Limitations and lessons learned

This study was a well-executed randomised design, with a MDES and level of attrition that meet the EEF's standards for high validity. The outcomes are all measured with established tools that have strong validity and reliability. In the IPE, good evidence has been provided to suggest that the intervention was implemented with high fidelity to the programme structure and resources. It has also been shown that, while concurrent interventions did take place during the trial, they

[42] https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit.

did so at very similar levels in the treatment and control groups. Nonetheless, the four most substantial trial limitations —outcome measurement, power, attrition, and sampling—are discussed below.

Outcome measurement

The primary outcome measure—GCSE Attainment 8—was collected over a year after the end of the intervention. This delay may explain the absence of a clear treatment effect. Over time, the initial benefits of the intervention could diminish or be influenced by external factors, such as changes in teaching quality, student motivation, or life circumstances. This may explain why we observe a significant effect on the survey outcomes, which were collected within months of the intervention ending, and not on the attainment outcomes. Another limitation relating to outcomes is that pupils' attitude to education (the key hypothesised mediating mechanism) was not measured quantitatively, limiting our ability to comment on whether the intervention had an effect in this regard. The decision was made with the EEF to only measure and estimate effects on changes that were classed as outcomes in the intervention's theory (changes that the developers considered to be important ends in themselves); to limit the resource spent on outcome measurement selection and design, and to limit the number of statistical tests conducted. This was perhaps a mistake; a quantitative estimate of the effect on pupils' attitude to education would have provided valuable information.

Power

The MDES of the randomised sample for the primary analysis was 0.19 SD; less than the EEF's threshold for a threat to validity, which is set at 0.2 SD (EEF, 2019, p. 6). However, the intervention was not designed to improve academic attainment, and the average effect of interventions evaluated by the EEF is much smaller than this (+0.04 SD; Demack *et al.,* 2021, p. 12). So, it is very likely that, if the Community Apprentice has any effect on attainment, it is much smaller than the 0.19 SDs that we were powered to detect. The study was therefore, likely unpowered for the primary outcome.

Attrition

From randomisation to analysis, 9% of observations were lost, either due to missing outcome data or missing covariate data. This is below the EEF threshold for a substantial threat to validity, which is set at 10% (EEF, 2019, p. 7). It is however, over the threshold that we pre-specified to consider imputation of covariates (set at 5%). When we explore this attrition, we find a slightly higher proportion of missing data in the treatment group as compared to the control group. Balance checks show no substantial differences between the two groups on observable characteristics in the analysed sample, but we cannot rule out some imbalance on unobservables that has been caused by this attrition. Any bias that is introduced into our estimated treatment effects will be at least partly addressed through covariate adjustment, but we have also carried out a sensitivity check where we remove all covariates from the model used to estimate the effect on the primary outcome. This sensitivity check reduces the attrition to 3% and the resulting estimated treatment effects are still within the 95% CI of the primary analysis.

Sampling

The sample for the impact evaluation differs from the general population in potentially important ways at the pupil and school levels. Pupils in the trial are much more likely to be entitled for FSM (47% vs 15%) and slightly more likely to be female (56% vs 49%). Schools in the trial are much more likely to be based in urban areas (100% vs 75%) and to be academies (60% vs 34%). However, the distribution of Ofsted ratings among trial schools was very similar to that of the general population of English schools. These differences in pupil-level and school-level characteristics mean that we cannot easily generalise the findings of this impact evaluation to the general English population. An increase in the proportion of rural schools in a nationally scaled version of the programme could make it more difficult to recruit high-quality coaches. This may in turn lead to a reduction in effectiveness.

The data collected from intervention schools as part of the IPE (either via survey or fieldwork visits) only represent the views and experiences of a subset of the treatment group. The qualitative findings are, therefore, not statistically representative, though the use of purposive sampling means that they should provide a good indication of the range and diversity of experiences and attitudes. Additionally, there may be some recall errors in survey responses.

Future research and publications

We have strong evidence from this study and the wider literature that youth social action can develop non-cognitive skills and attributes. There is also promising large-scale observational evidence that similar non-cognitive skills and

attributes can lead to improved labour market outcomes (Gutman and Schoon, 2013; Heckman and Kautz, 2013). The next stage is to causally test whether this chain holds, exploring whether participation in youth social action can lead to long-term, material benefits for its participants. This provides a strong justification to carry out the longitudinal analysis that has been provisionally planned for the Community Apprentice, to find out if the programme has a positive effect on employment status post-18 years old (or a similar labour market outcome). It may also be worth estimating effects post-16 educational attainment and participation in higher education (the other two provisional longitudinal outcomes). However, the null result on GCSE attainment suggests that any effect on labour market outcomes is more likely to be mediated by the non-cognitive skill development, rather than academic progress.

References

Anders, J. (2016) 'Department for Education Longitudinal Study of Young People in England'. (Unpublished).

Birdwell, J., Scott, R. and Reynolds, L. (2015) 'Service Nation 2020'. London, UK: Demos.

Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), pp.77-101.

Cameron, D. Stanndard, J., Leckey, C., Hale, C., & Di Antonio, E. (2017) 'National Citizen Service 2015 Evaluation, Main Report'. London, UK: Department for Culture, Media & Sport.

Chen, G., Gully, S.M. and Eden, D. (2001) 'Validation of a New General Self-Efficacy Scale'. Organizational Research Methods, 4: 1, 62–83.

Collaborative for Academic, Social, and Emotional Learning. (2015). Effective Social and Emotional Learning Programs: Middle and High School Edition. CASEL.

Corcoran, R. P., Cheung, A. C., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. Educational Research Review, 25, 56-72.

Demack, S., Maxwell, B., Coldwell, M., Stevens, A., Wolstenholme, C., Reaney-Wood, S., Stiell, B. and Lortie-Forgues, H. (2021) 'Review of EEF Projects. Evaluation Report'. London, UK: Education Endowment Foundation.

Department for Education. (2024) Secondary accountability measures: Guide for maintained secondary schools, academies and free schools. Department for Education.

Education Endowment Foundation. (2019). Classification of the security of findings from EEF evaluations. Education Endowment Foundation.

Fitzpatrick, A., Matthews, P., Hassan, A., Greene, O., Bates, J., Miller, J. and Conlon, G. (2021) 'National Citizen Service 2019 Summer Evaluation: Main Report'. London, UK: Department for Digital, Culture, Media & Sport.

Gorard, S., See, B.H., Siddiqui, N., Smith, E. and White, P. (2016) 'Youth Social Action Trials: Youth United. Evaluation Report and Executive Summary'. London, UK: Education Endowment Foundation.

Gutman, L.M. and Vorhaus, J. (2012) 'The Impact of Pupil Behaviour and Wellbeing on Educational Outcomes'. London, UK: Department for Education.

Gutman, L.M. and Schoon, I. (2013) 'The Impact of Non-Cognitive Skills on Outcomes for Young People. A Literature Review'. London, UK: Education Endowment Foundation.

Heckman, J.J. and Kautz, T. (2013) 'Fostering and Measuring Skills: Interventions That Improve Character and Cognition'. NBER Working Paper Series. Cambridge, MA: National Bureau of Economic Research (NBER).

Humphrey, N., Barlow, A., Wigelsworth, M., Lendrum, A., Pert, K., Joyce, C., Stephens, E., Wo, L., Squires, G., Woods, K., Calam, R., Harrison, M., & Turner, A. (2015). Promoting Alternative Thinking Strategies (PATHS): Evaluation report and Executive summary. Education Endowment Foundation.

Humphrey, N., Hennessey, A., Ashworth, E., Frearson, K., Black, L., Petersen, K., Wo, L., Panayiotou, M., Lendrum, A., Wigelsworth, M., Birchinall, L., Squires, G., & Pampaka, M. (2018). Good Behaviour Game. Evaluation Report and Executive Summary. University of Manchester. Education Endowment Foundation.

Kerr, D. (2014) 'We Need Renewed Political Commitment to Citizenship Education and Ongoing Monitoring of Its Provision in Schools'. London, UK: Democratic Audit.

Kirkman, E., Sanders, M., Emanuel, N. and Larkin, C. (2016) 'Evaluating Youth Social Action: Does Participating in Social Action Boost the Skills Young People Need to Succeed in Adult Life?' London, UK: The Behavioural Insights Team.

Lower-Hoppe, L.M., Newman, T.J. and Anderson-Butcher, D. (2017) 'Validity and Reliability of the Teamwork Scale for Youth'. *Research on Social Work Practice,* 27: 6, 716–25.

McCroskey, J.C. and McCroskey, L.L. (1988) 'Self-Report as an Approach to Measuring Communication Competence'. *Communication Research Reports,* 5: 2, 108–13.

National Audit Office. (2017) 'National Citizen Service. Report by the Comptroller and Auditor General'. London, UK: Cabinet Office and Department for Culture, Media & Sport.

Pye, J. and Michelmore, O. (2017) 'National Youth Social Action Survey 2016'. London, UK: Ipsos MORI.

Ritchie, J., Lewis, J., Nicholls, C.M. and Ormston, R. eds., 2014. Qualitative research practice: A guide for social science students and researchers. Sage.

Saldana, J., 2009. The Coding Manual for Qualitative Researchers. Sage.

Schulz, K.F. and Grimes, D.A. (2002) 'Sample Size Slippages in Randomised Trials: Exclusions and the Lost and Wayward'. *Lancet,* 2: 359, 781–85.

Spera, C., Ghertner, R., Nerino, A. and DiTommaso, A. (2013) 'Volunteering as a Pathway to Employment: Does Volunteering Increase Odds of Finding a Job for the Out of Work?' Washington, DC: Corporation for National and Community Service, Office of Research and Evaluation.

Sylva, K. Melhuish, E., Sammons, P., Siraj-Blatchford, I., Taggart, B., Toth, K., Smees, R., Draghici, D., Mayo, A. and Welcomme W. (2012) 'Effective Pre-school, Primary and Secondary Education 3–14 Project (EPPSE 3–14) Report from the Key Stage 3 Phase: Influences on Students' Development From Age 11–14'. London, UK: Department for Education.

Taylor, P.J. (2021) 'Service Learning and Youth Political Participation: A Mixed-Method Thesis, Doctoral Thesis'. London, UK: University College London.Teddlie, C. and Tashakkori, A. (2009) Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioural Sciences. Sage, London.

Wigelsworth, M., Squires, G., Birchinall, E., Kalambouka, A., Lendrum, A., Black, L., Troncoso, P., Santos, J., Ashworth, E., & Britteon, P. (2018). FRIENDS for life: Evaluation report and executive summary. Education Endowment Foundation.

Zimenkova, T. (2013). Sharing Political Power or Caring for the Public Good? The Impact of Service Learning on Civic and Political Participation. In Hedtke, R. & Zimenkova, T., (Eds.), Education for Civic and Political Participation: A Critical Approach. Oxon: Routledge: pp. 171-188.

Appendix A: EEF cost rating

Table 33: Cost rating

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

Appendix: Padlock rating

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	MDES	Attrition			Adjustment for threats to internal validity [-1]		
5	Randomised design	<= 0.2 (=0.19)	0-10% (=9%)	5				
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					4
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%					
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Moderate	RCT design to control for observables and unobservable differences and allocation to trial arms conducted independently. Although individually randomised, there is no perceived risk of contamination. Baseline control measures used in the analysis: Math and Reading KS2 scores. The imbalance in baselines between treatment and control (as randomised) was 0.074 for Math and 0.078 for reading (Hedges' g). Since differences are greater than 0.05, the threat has been classified as moderate.
Threat 2: Concurrent Interventions	Low	Existence of extracurricular activities that could have had an impact in primary and secondary outcomes were explored by the evaluation team. There was no significant difference between the intervention and control groups in terms of their participation in other extracurricular activities. Both peer reviewers classified this risk as low.
Threat 3: Experimental effects	Moderate	One peer reviewer rated this risk as low but was unsure and pointed at the comparative number of concurrent interventions in BAU group, and possible compensatory rivalry (depending on when interventions were allocated). The second peer reviewer classified this risk as moderate due to the use of attainment as primary outcome, which measure-underpowered the trial.
Threat 4: Implementation fidelity	Moderate	Fidelity was defined and examined by the research team. However, coach quality was shown to be important through the IPE and appeared variable. It may be that fidelity was lower in quality, thus washing out the programme effects. One peer reviewer classified this risk as moderate due to the coaching variation. The other one as low but they were unsure.

Threat 5: Missing Data	Low	9% overall missing data for the primary outcome, higher for the treatment group (12% vs 7%). The analysis does a good job of dealing with missing data, and there are no substantial differences between the two groups in terms of observable characteristics. Both peer reviewers classified this risk as low.
Threat 6: Measurement of Outcomes	Low	A statutory measure (GCSE) was used and no ceiling or floor effects were reported. One peer reviewer classified this risk as low/moderate, arguing that a year time lapse between end of the intervention and students sitting GCSEs could have washed out of any effect of the intervention. The other peer reviewer classified this threat as low.
Threat 7: Selective reporting	Low	Analyses reported clearly and per SAP.

- **Initial padlock score:** 5 Padlocks - A solid RCT reasonably powered and with moderately low attrition.
- **Reason for adjustment for threats to validity**: 1 Padlock – Some threats to the internal validity of the trial could have diluted some effect of the intervention on the primary outcome.
- **Final padlock score:** 4 Padlocks.

Appendix C: Effect size estimation

Table 34: Effect size estimation

			Intervention group		Control group			
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance	Population variance (if applicable)
Attainment 8 score	-0.87	0.124	345 (45)	330.21	344 (27)	345.16	337.68	NA
English attainment	-0.24	-0.073	345 (45)	13.69	347(24)	14.53	14.11	NA
Maths attainment	-0.257	-0.085	347(43)	17.13	345 (26)	16.93	17.03	NA
Teamwork	0.05	0.285	332 (58)	0.257	333 (38)	0.252	0.254	NA
Self-efficacy	0.10	0.108	311 (79)	0.309	312 (59)	0.316	0.313	NA
Social confidence	5.49	5.17	333 (57)	345.34	328 (43)	300.82	323.25	NA

NA, not applicable.

Further appendices

Further appendices can be found on the EEF website.

This document is available for download at https://educationendowmentfoundation.org.uk

Education Endowment Foundation

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

https://educationendowmentfoundation.org.uk
@EducEndowFoundn
Facebook.com/EducEndowFoundn