THE
BEHAVIOURAL
INSIGHTS TEAM

Behavioural Insights Team

Evaluating Youth Social Action

Does Participating in Social Action Boost the Skills Young People Need to Succeed in Adult Life?

Elspeth Kirkman
Michael Sanders
Natalia Emanuel
Chris Larkin

THE BEHAVIOURAL INSIGHTS TEAM

Contents

Social action: A definition

Social action can broadly be defined as 'practical action in the service of others to create positive change'. This Step Up To Serve definition covers a wide range of activities that help other people or the environment, such as fundraising, campaigning (excluding political campaigning), tutoring/mentoring, and giving time to charity.

Executive summary

This report provides compelling and robust evidence that young people who take part in social action initiatives develop some of the most critical skills for employment and adulthood in the process. Using rigorous evaluation techniques and drawing on a mature field of research linking specific character skills to hard outcomes, such as employability, we are able to demonstrate that – for those programmes we have evaluated – investment in social action in general leads to benefits for young people taking part as well as for the intended beneficiaries.

In 2013, The Cabinet Office Centre for Social Action and the Education Endowment Foundation invited applications from organisations working with young people in Birmingham, Kent, Middlesbrough and Lancashire to receive grants from a £5 million Youth Social Action Fund. This invitation was part of the Cabinet Office's support to the Step Up To Serve #iwill initiative, a cross-party campaign championed by HRH The Prince of Wales with a mission to improve the quality, quantity and frequency of social action for young people aged 10-20 by 2020. 28 organisations were given funding through the programme, working across the country on a diverse range of social action projects targeted at young people in a variety of settings.

The Fund was also envisaged as an opportunity to capture high quality data on the extent to which participating in social action initiatives helps to prepare young people for adulthood and, specifically, work. To this end, the Behavioural Insights Team was commissioned to conduct the evaluation of the programme, with a focus on key character skills for adulthood and employment. This report focuses on the evaluation aspect of the programme.

Historically, evaluations of the effects of social action on young people's broader prospects have been qualitative in method. The ambition for this programme was to add to this existing research-base by providing quantitative evidence to identify whether social action in and of itself causes and catalyses the development of skills for work and life in young people who take part. To achieve this, we ran

three randomised controlled trials (RCTs) and one pre/post comparison to compare the outcomes for young people who took part in these funded initiatives against the outcomes of young people who did not. This report builds on the earlier, interim report, through the addition of the pre-post analysis. We also conducted further trials to test whether the survey responses young people gave translated into differences in actual behaviour. In a further addition to the interim report, this includes a simulated interview task that captures further measures of employability. To capture the breadth and impact of activity across the other Providers, we drew on the data collected by the Social Action Passport; a digital tool used across the country by young people participating in the various funded initiatives. This analysis will be included in a later release of this report.

In our randomised controlled trials, we find that those who participated in the programmes display significant improvements to their skills for work and life compared to their counterparts, as measured by reliable and validated questions. We also find that young people who took part in these social action initiatives are more likely to express interest in specific volunteering activities available to them in the future than their counterparts who have not engaged in a social action initiative. Conversely, we also find a decreased willingness to donate money to charity when the same group is given discretion to use a small amount of money as they wish; a finding that brings to light interesting questions about the distinction between giving and volunteering and the ways in which young people view the relative contribution of different activities. Finally, some participants in the treatment and control groups of one study were asked to complete a short interview, which we have used to assess whether their interview performance has been improved by the treatment. We find that social action participation improves performance on the interview task, although this is not statistically significant. In the programme not evaluated using a randomised trial, the small sample and lack of robust counterfactual group mean that although the majority of findings are positive, statistical significance is not achieved.

This report describes the way in which the research methods and measures were chosen, the programmes themselves and the results. This analysis is extremely encouraging for both the youth sector, which has historically struggled to quantify its value-added in a way that can help funders make tough decisions, and for government, whose investment in youth social action is helping young people to realise their potential and become the skilled adult citizens of tomorrow.

Background and context

In 2013, The Cabinet Office Centre for Social Action and the Education Endowment Foundation (EEF) invited applications from organisations working with young people in Birmingham, Kent, Middlesbrough and Lancashire to receive grants from a £5 million Youth Social Action Fund. The Centre for Social Action aims to identify and accelerate the development and spread of high impact social action initiatives by:

1. Identifying social action innovators.
2. Supporting organisations with promising initiatives to scale up and become sustainable.
3. Developing the evidence base on the impact of social action-based interventions.
4. Working with policy makers and those delivering public services to help them understand the contribution that individuals and communities can make.
5. Mobilising large numbers of volunteers by making the opportunities to volunteer appealing and the act easy.
6. Supporting the development of a modern and effective infrastructure to support social action.

This invitation for applications was part of the Cabinet Office's support to Step Up To Serve. Step Up To Serve runs the #iwill campaign which aims to increase the number of young people aged 10-20 taking part in social action by 50 percent by 2020. The campaign was launched in November 2013 by HRH The Prince of Wales, with cross-party support. As well as increasing the number of young people engaged in social action, the campaign also aims to increase the quality of programmes, through its six principles of youth social action. To this end, the remit of the Fund was to offer financial support to programmes which:

- Provide social action opportunities for 10-20 year olds, focusing either on 10-14 year olds, or demonstrating how they fill a gap in existing provision for other age groups within the 10-20 age range; or
- Support the take-up of, and progression between, social action opportunities for 10-20 year olds, and the overall success of those programmes.

- Address Step Up To Serve's six principles of social action:

    1. Reflective
    2. Challenging
    3. Youth-led
    4. Socially Impactful
    5. Progressive
    6. Embedded

In addition to this direct call for Providers of social action, the Fund sought applications from organisations that could support the work by providing:

- A Social Action Passport that would enable young people to record all of their social action outcomes in one place, and
- An award system that would recognise and reward the social action of young people, and create a useful shorthand for young people to describe their social action.

Funds were given out in late 2013 to 28 providers running a diverse range of programmes of varying sizes. Beat Bullying were commissioned to build and manage the Passport and vInspired to run the awards. The Behavioural Insights Team was appointed as the evaluator for the 28 providers. EEF are carrying out separate randomised controlled trials on two other programmes, to establish the link between social action and educational attainment. Results from these evaluations will be published separately.

The next section of this report discusses the process and rationale for choosing the specific research methods used on the programme.

The research methods used

Aside from simply boosting labour market participation, the Cabinet Office was interested in using the Fund as an opportunity to gather rigorous evidence in relation to the questions of whether social action can help its young participants to develop critical skills for employment and adulthood.

To achieve this, it was decided that randomised controlled trials would be used to determine programme impact. This section introduces the concept of RCTs as well as the rationale for choosing the programmes that have been evaluated in this way. The *About the measures used* section of the report discusses how we determined what to measure and the way in which we did it to capture outcomes.

About randomised controlled trials

Since its inception in 2010, The Behavioural Insights Team has run more randomised controlled trials than had ever been run before in the history of government. Members of the team along with leading academics co-authored the seminal report "Test, Learn, Adapt"[1], which was released through the Cabinet Office in 2012 and was instrumental in kick-starting a government-wide move towards using RCTs to build and assess public policy.

The RCT method uses similar techniques to those deployed in testing the effectiveness of new medicines. By taking a large enough group of trial participants (in this case, young people in schools), it is possible to allocate participants such that some receive the "treatment" (in this case, participation in a social action initiative) and some do not. Those that do not are referred to as the "control" group. Because the participants are randomly allocated into the two groups, there are no systematic differences between those who receive the treatment and those who do not; in other words, they are, on average, the same across every dimension on which individuals might differ. Thus we can conclude that any differences between the two groups observed after the treatment were caused by the treatment rather than by some other factor.

This is an enormous advantage when working with RCTs; they allow us to make causal inferences. As already described, causal inference relies on having a high degree of confidence that those in the treatment and control groups are sufficiently similar that the control group can provide a "counterfactual" – a simulation of what would have happened to those who received the treatment if

[1] https://www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials

they had not received it. RCTs allow for strong conclusions to be drawn about the efficacy of specific interventions in relation to specific outcomes of interest.

There are many practical examples that highlight the worth of RCTs. For example, "Scared Straight" is a programme developed in the US to deter juvenile delinquents and at-risk children from criminal behaviour. The programme is based on an intuitive and sensible theory: children would be less likely to engage in criminal behaviour if they were made aware of the serious consequences of crime through visits to criminals in custody. Several early studies, which used simplistic evaluation models that looked at the criminal behaviours of participants before and after the programme, seemed to support this idea[2] with reported success rates as high as 94%. The programme was adopted in several countries, including the UK. Problematically, none of these evaluations had a control group showing what would have happened to these participants if they had not participated in the programme. Several RCTs set out to rectify this problem. The RCTs revealed something surprising: "Scared Straight" in fact led to higher rates of offending behaviour. In other words, "doing nothing would have been better than exposing juveniles to the program"[3] and taxpayers were spending significant amounts of money on an initiative that actively increased crime. Not only does this instance underscore the importance of RCTs as an evaluation mechanism, it is a helpful reminder that, in many cases, it is unclear what works without rigorous testing and even the most intuitive of ideas may not translate into the outcomes we expect.

For our trials, half of the participating schools and, by extension, their eligible students got the opportunity to participate in a specific programme under the Youth Social Action Fund whilst the other half served as our control group. We used a "pipeline design", meaning that the schools assigned to the control condition would also get the opportunity to participate in the programme once we had observed the effect of the programme on the treatment group of schools. This was a necessary feature of the design to ensure that the control group schools were sufficiently incentivised to participate in the trial.

Aside from the ask of schools, it was recognised that running trials of this nature as part of the programmes would be challenging, requiring extra work from the Providers (for example, administering surveys to both treatment and control schools and recruiting from a pre-specified pool of schools to get a diverse

[2] Finckenauer J. O. (1982) Scared Straight and the Panacea Phenomenon. Englewood Cliffs, NJ: Prentice-Hall, 1982.
[3] Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2003). Scared Straight and other juvenile awareness programs for preventing juvenile delinquency. Campbell Review Update I. The Campbell Collaboration Reviews of Intervention and Policy Evaluations (C2-RIPE). Philadelphia, Pennsylvania: Campbell Collaboration.

range). When running RCTs, sample size also matters. We needed to ensure we had a large enough sample of young people for each Provider programme to be certain of balanced treatment and control groups, thereby ensuring our ability to detect differences between them with certainty. With these requirements (effort and size) in mind, we chose three providers – Citizenship Foundation, Envision and Voluntary Action within Kent – that were suitable for RCT evaluations and one, CSV, which could make use of a matched pairs design. More detail on each programme is given in the next section.

Once the four Providers had been selected, we worked together to understand how each of the projects would function and what the likely effect of taking part would be on participants. By doing this, we were able to identify some common measures of impact that could be used across the four projects. These measures are described in the *About the measures used* section of this report and informed the design of the research tools used. Because we could only assign young people to receive either the treatment or control, and not compel them to do so, the analysis of the trial results should be thought of as intention-to-treat analysis. While this slightly decreases the confidence with which we can interpret the effect of the treatment on the outcome, it does provide the added benefit of increasing external validity – and in turn, scalability.

The other aspects of the evaluation

As mentioned previously, for CSV we deploy a pre/post design. In some instances, randomisation is not feasible – often due to the smaller scale of a programme in terms of the numbers of schools involved. In such instances, a matched pairs design is often an appropriate evaluative tool, and it was our intention to implement this strategy. In a matched pairs design, instead of using random allocation, schools are systematically matched with a pair based on the features we can measure readily. Students in one school participate in the programme whilst those in the other do not. The resulting outcomes for both sets of students are then compared in the same way as in the randomised controlled trials. Unfortunately, Matched Pairs was not viable in this sample as the treatment sample had few observations drawn from a single school, making matching unreliable. As such, we compare participants post-survey responses to their own pre-survey responses for the 22 participants for which these matches are available. Of course, there may be external factors which also influenced our outcomes, not captured by these measures. These unmeasured factors lessen the confidence with which we can make causal inferences since we cannot say with certainty that there are no systematic differences influencing the participants other than the

programme. Because of these reasons the CSV results are reported in the *Technical Appendix*.

For the remaining 24 Providers – those that are not being evaluated using the RCT and pre/post designs – a more pragmatic approach was taken. To evaluate activity across a range of programmes, draw out themes in terms of what was effective and valuable for the participants, and capture the breadth of activity in each initiative, we are analysing the data captured from participants in the online Passport. This aspect of the evaluation will not be completed until the final programme ends and is not covered in this version of the report.

The next section describes the content of the four projects chosen for RCT and pre/post evaluation.

About the programmes being evaluated

The four programmes receiving in-depth evaluations, chosen because of their size, location, setting and subject matter, are as follows:

Citizenship Foundation

Founded in 1989 as a way of teaching young people about what is entailed in citizenship, the Citizenship Foundation (CF) has since expanded to include other programmes including those targeted at teachers, those lobbying for policy improvements, and those inspiring youth action.

Go-Givers is a cross-curricular PSHE and citizenship programme for primary schools with over 60,000 registered teachers accessing its materials. The Make a Difference Challenge is a pupil-led project for KS2 that supports pupils in identifying, researching and addressing a cause that they want to do something about. By developing a campaign, fundraising or taking direct action pupils develop awareness and skills for social action. Last year around 10,000 pupils in 190 schools took part in the Challenge.

Community Service Volunteers

Community Service Volunteers (CSV) aims to enable people of all ages to take an active role in their communities, whether by volunteering or by mentoring young people, enabling disabled people to lead independent lives, or offering security to families and children at risk of neglect.

As part of the Youth Social Action programme, CSV aimed to include an additional 380 young people across Dover and Canterbury in social action initiatives.

Envision

Founded in 2000, Envision is an organization dedicated to "enabling young people to become more aware of social and environmental problems and empowered with the self-belief and skills they need to build a better world". With locations in Bristol, Birmingham and London, Envision works with 130 schools. In their most recent academic year, they included 2,000 young people in their social action initiatives.

Social action programmes with Envision tend to be 10-months long and geared toward 16-19 year olds. Programmes are designed by young people to address their own local community's needs such as race relations or knife violence. Envision aims to increase self-efficacy, accountability, and the ability to be a proactive problem-solver by letting young people lead the programmes.

Voluntary Action Within Kent (VAWK/IMAGO)

Voluntary Action Within Kent (VAWK) aims to support a range of populations within Kent from those with disabilities to elders with particular needs to young people.

VAWK's Youth Social Action programme operates in 25 schools across Kent, engaging over 5,000 young people. Participants are encouraged to take the lead and develop the projects themselves with support provided by mentors at VAWK and in their schools. In addition, the programme is aimed at having additional sustainability because specially selected 15-18 year olds help lead the programme by including and encouraging the participation of younger students.

At the time of writing, VAWK is rebranding as IMAGO and will be referred to throughout the remainder of this report as VAWK/IMAGO for posterity.

As stated in the previous section, common outcome measures were determined for each provider project and were used to inform the way in which we collected information from participants. The next section describes how we collected this data and why we chose the measures and methods deployed.

About the measures used

For the purposes of this work, the Cabinet Office set a focus on measuring changes in those soft skills that are shown, through extensive research[4], to prepare young people for successful adult life and, notably, positive employment outcomes. Much groundwork had already been done to translate this into the youth services domain, including the publication of the Quality Framework for

[4] For a good overview, see Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. Education Endowment Foundation. Available at: http://educationendowmentfoundation.org.uk/uploads/pdf/Non-cognitive_skills_ literature_review.pdf.

Youth Social Action[5], which provided a starting point for devising the measures for these trials. This framework is comprised of a set of measures designed to assess a range of community benefits, from civic participation, health, educational engagement to safer communities, sustainability, voting, resilience and, crucially, employability.

To sharpen the focus on employability, we identified six key constructs – characteristics that are not directly observable in measures such as attainment but that resonated with the objectives of each of the four programmes' goals and are shown through previous research to link to positive life outcomes, such as employability[6]. The constructs are as follows:

Construct	Definition
Empathy	The ability to understand and share the feelings of another.
Problem Solving	The ability to reason, use available information and think laterally in order to reach a goal or end point.
Cooperation	Working together with others to the same end or goal.
Grit and resilience	Grit is the tendency to sustain interest in and effort toward very-long term goals. Resilience is the ability to bounce back.
Sense of Community	Identification as part of a community, perception of agency within it and propensity to take prosocial action.
Educational Attitudes	Understanding of the value of education and taking an interest in building knowledge and skills.

We also measured levels of interest in future volunteering opportunities, wellbeing levels and, for VAWK/IMAGO and Citizenship Foundation, social trust scores (which draw out how much the participants think other people can be trusted).

All of the measures in the survey, as well as the broader premise of measuring soft skills as the primary outcome, were chosen from a long list of possible options based on the following:

1. Relationship to employability

[5]http://youngfoundation.org/wp-content/uploads/2013/08/Scoping-a-Quality-Framework-for-Youth-Social-Action-FINAL.pdf
[6] See validated survey questions in the appendix which have been designed drawing on this literature

2. Malleability of the construct
3. Extent to which the programmes aimed to influence this construct
4. The quality of existing validated[7] tools to measure these constructs (and hence comparability to other studies)

All four of these factors were important to consider. For example, employment might be most strongly correlated with student grades, and can be readily validated. However, Sanders & Ní Chonaire (2015) find that this measure is incredibly difficult to move in general, and particularly with tangentially related interventions. At the same time, Youth Social Action programmes were not developed with the intention of improving this outcome, although we do seek to explore impact on this measure in later iterations of this analysis. By contrast, social skills, also correlate with employability and can be demonstrated through measures of students' empathy and cooperation, using reliable and valid survey questions. These outcomes are easier to capture and are also more malleable in the short term. In addition they overlap directly with the stated goals of the Providers, who explicitly identify boosting these attributes as an objective. Other measures, such as the social trust questions, are common across both the Civil Service employee wellbeing survey, and the evaluations of similar programmes such as National Citizen Service (NCS). Measures were chosen from a long list of possible metrics in collaboration with the Providers themselves, whose aims were critical in shaping what we measure, with the Cabinet Office and with others in the youth sector.

With these things in mind, it was important to ensure that we had valid and reliable surveys. That is to say, questions had to genuinely measure the trait intended to be assessed and be capable of producing the same results when repeatedly measuring the same group. Moreover, questions ought not to measure something else, such as the respondent's eagerness to please the survey administrator. Surveys that have not been validated are likely to have pitfalls. For example, one question that had previously been used to review an intervention asked young people to indicate whether they agree or disagree with the statement "I have a greater empathy for other people." Such a question is problematic because many young people may just want to please the survey administrator or indicate their own emotional awareness—regardless of whether their perspective has been enhanced by the programme.

[7] Validation is the process by which research instruments are confirmed to be accurately measuring that which is set out to be measured. This process also tests reliability; confirmation that retaking the measure again at a different time under the same conditions will yield the same result.

In validating survey questions, considerable effort goes into testing, refining questions, and cross-referencing answers to the actions and outcomes of those who answer them. For instance, it can be important to correlate self-reports of empathy with actual giving behaviour, and to relate reports of skills gained to actual employment status or tasks that require said skills. For this reason, we relied on standard surveys that are used by researchers to measure specific constructs, traits, or skills. Reliability of the answers received is also a key point to consider. Questions are reliable if they elicit the same response in the same context time and again. If questions are vague, open to interpretation, complex or hard to understand in some other way, then the answers may not produce reliable data.

All of this means that the ideal approach is to reuse existing validated survey questions in their totality, something we were able to do for the wellbeing questions and the question on social trust. However, for the six constructs, there were complexities in relation to this specific piece of work. For example, in survey-administration, the more questions asked, the more likely it is that the survey will draw a reliable and valid conclusion from the responses given because a plethora of questions can encompass all aspects of any given construct in a way that fewer questions cannot. As a result, many validated surveys that are routinely used by researchers to measure specific traits, skills or constructs are lengthy and – given the age of our participants and the time they had available – consolidating validated questions for a six-construct assessment was unrealistic.

As a result, the decision was taken to choose appropriate questions from a range of validated surveys and create a new amalgamated survey designed specifically to measure the impact of youth social action in relation to the six constructs. The questions chosen to test these constructs are included in the *Instrument* section of the Technical Appendix and the full survey can be found in the *Survey Questions* section. All questions were agreed on in conjunction with a wider group of stakeholders, including the Education Endowment Foundation. The validity of our survey was tested once we received the responses by running statistical tests to see whether the answers correlate with one another in the same way as those in the longer, already validated, surveys. More detail is provided in *the results* section of this report where we find that all of our construct measures, except for "educational attitudes", meet the required standard. As such, the results regarding educational attitudes should be regarded with somewhat less confidence than the results of the other measures, although overall the survey can be reused.

In addition to these survey measures, we are also collecting data on participating students from the National Pupil Database to see whether the programmes had an effect on 'hard' outcomes such as attainment (where key stage exams are taken) and attendance at school, which predicts attainment later on. This aspect of the work has not yet been completed and will be covered in a later report.

To supplement the surveys, we also sought to measure observable behaviour in two tasks conducted with sub-groups of participants from both treatment and control schools. As with the surveys questions, we endeavoured to use tasks that have been widely used in previous research. These are as follows:

1. **A donation opportunity:** young people from the treatment and control groups in the Citizenship Foundation programme were given four 50 pence pieces and told to decide in private (by placing the money in whatever proportions they saw fit in padded envelopes marked "me" and "charity") whether they would like to keep it for themselves or donate it to The British Red Cross. Four fifty pence pieces (£2) was chosen as an optimal amount as it was enough to buy something attractive and readily available, sweets, for example, as an alternative to donating. This technique for learning about people's social preferences is widely used in experimental economics[8].
2. **An interview task:** Envision students from both the treatment and control groups were filmed in a mock job interview. Their performance was assessed in two ways. Firstly, using an objective scoring system that was created in consultation with a range of companies who recruit young people out of school. Secondly, experienced hirers assessed whether they would hire the young person or not based on a "gut reaction". This mechanism has been used in other studies of employability[9].

The rationale behind including these two behavioural simulations is that they may be better at revealing preferences that may not otherwise be well captured. In the first instance, while a student may feel that he or she is more or less likely to donate their time or money to charity after the programme, it is important to understand what they actually do when they have the opportunity. Similarly, in the latter simulation, understanding how the social action venture has actually had an impact on employer's willingness to hire is essential towards understanding if the programme has actually had an effect on this outcome measure.

[8] E.g. See Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The journal of economic perspectives*, 153-174.
[9] E.g. Cuddy, A. J., Wilmuth, C. A., & Carney, D. R. (2012). The benefit of power posing before a high-stakes social evaluation. And (for broader behavioural predictions) Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, *111*(2), 256.

The results

In this section, we discuss the trials for which we already have results in relation to the following areas:

Validation of survey measures
Survey-based trials:

- Overall results for the three RCTs from surveys by programme
- Expressions of willingness to volunteer in the future

Behavioural trial:

- Willingness of Citizenship Foundation participants (treatment and control) to donate to charity

Validation of survey measures

As discussed in previous sections, the need to design bespoke surveys that would meet the needs of our participants meant that the first task in getting results was to validate the surveys to check they measured the things we wanted them to capture in a consistent, reliable, fashion.

We find that, the majority of our measures are valid and reliable with all but our "educational attitudes" measure meeting the standard set for reliability (see *Technical Appendix* for more information). As such, we are confident that the results are not only measuring what they intend to but that the surveys could be used again to measure outcomes by the sector. Rather less confidence may be attributed to the results on students' educational attitudes.

Survey-based trials

Overall results from pre and post surveys by Programme

To recap, the method of evaluation for the three trials reported in this document involved:

- Each of the Providers (Citizenship Foundation, Envision and VAWK/IMAGO) recruiting schools;
- Random allocation of each school (and, therefore, its pupils) to receive treatment – that is to say, the chance to undertake the Providers' social action programme – or control (offer of the programme but starting at a later point in time);
- Students in the eligible age range in all schools (treatment and control) answering a set of questions in a pre-survey designed to elicit an understanding of their skills relative to each of the constructs we are testing for. This allows us to establish a baseline and to confirm that random allocation was successful (i.e. that the students are comparable from the outset between treatment and control schools);
- The Provider running their programme in treatment schools;

- Students in both treatment and control schools taking the same survey again at the end of the programme so we can compare the scores at a second point in time to see what the effect of the Provider programme was on participants.

For each Provider, we see that the random allocation was a success as there are no significant differences between the pre-survey scores for treatment and control schools. We also see a positive and significant impact in the post-surveys as a result of taking part in the initiatives in the scores of our treatment students. The results are displayed for each programme below[10].

Note that scores for the constructs are standardised on a 10-point scale. This is because each of the constructs we measure had a different number of questions associated with them and, therefore, a different total possible score. By presenting on a standardised scale, we remove the distraction this creates when comparing the baseline and changes in each construct. Moreover, this standardisation allows readers to explicitly compare the sizes of the effects.

Citizenship Foundation results

The programme was very effective in increasing empathy levels, problem solving, grit, and community skills relative to control students – in each of these areas, we see statistically significant differences. Note that analysis is conducted at this stage on the population for whom all outcome measures are available. Young people who experienced the programme showed, on average, a level of empathy 6% greater than those who did not experience the intervention. Those students were also more adept in problem-solving than the control students. Additionally, those who participated in social action showed a level of grit (7.0) that was significantly above that of the young people who did not participate (6.4). Similarly, the level of community investment amongst young people was considerably higher amongst participants than control students. We saw no statistical differences between the control and treatment groups on attitudes to education or cooperation.

[10] Note: the use of *, ** and *** is a statistical convention. These denote statistically significant results, if there is no * then difference between treatment and control is not deemed significantly different, although if we had a larger sample, the same amount of difference *could* eventually become significant. The more *s, the higher the confidence that the result is down to the Providers' intervention (* = 95% confidence, ** = 99% confidence, *** = 99.9% confidence).

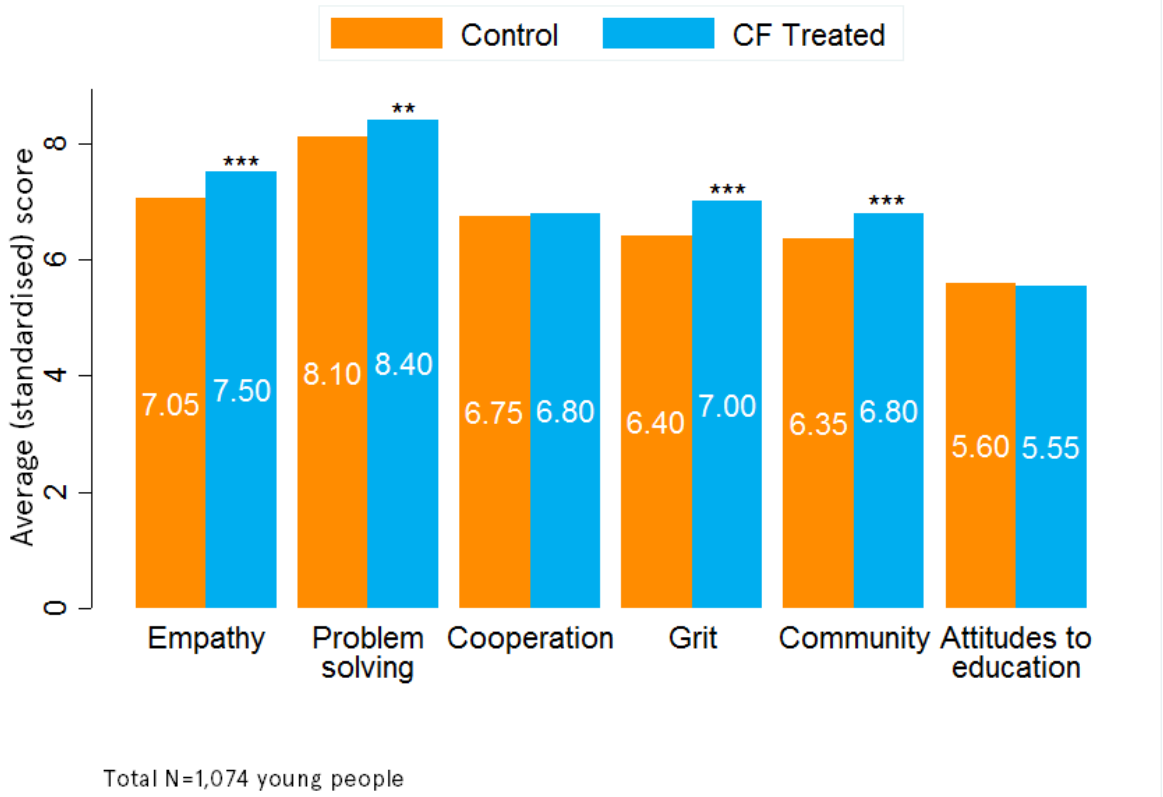


Figure 1: Citizenship Foundation results for the six psychological constructs

We also find that those who take part in the programme have a more positive outlook; stating that things in life are worthwhile more often than their peers and also reported lower levels of anxiety (a decrease of 22%).

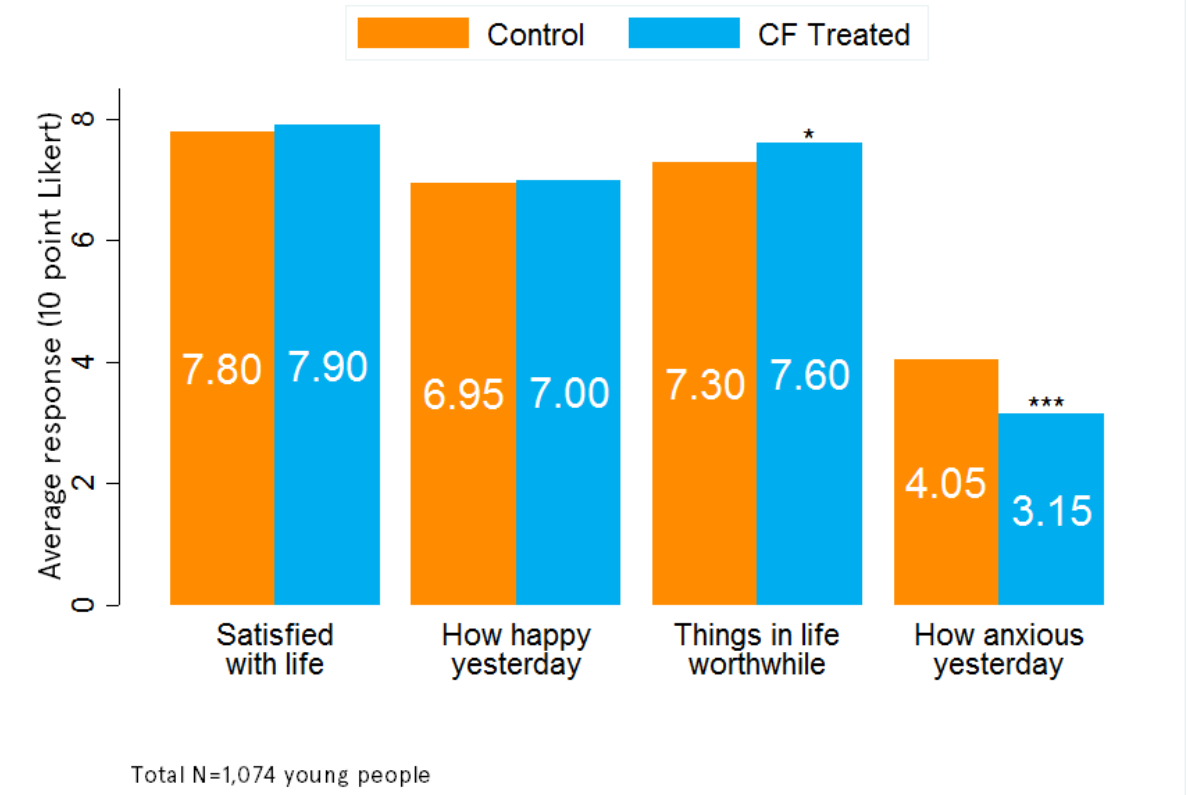


Figure 2: Citizenship Foundation results on wellbeing levels

Finally, we measure social trust by asking how much other people can be trusted. The possible answers are coded 1, 2, 3 and 4 for the purposes of data analysis and are as follows: 1 = No one can be trusted; 2 = Some people can be trusted; 3 = Most people can be trusted; 4 = Everyone can be trusted. Although social trust is important in adult life, it is perhaps unsurprising that we do not see a marked change in a population of comparatively young children who are likely to receive strong and frequent messages about trust (e.g. "don't talk to strangers").

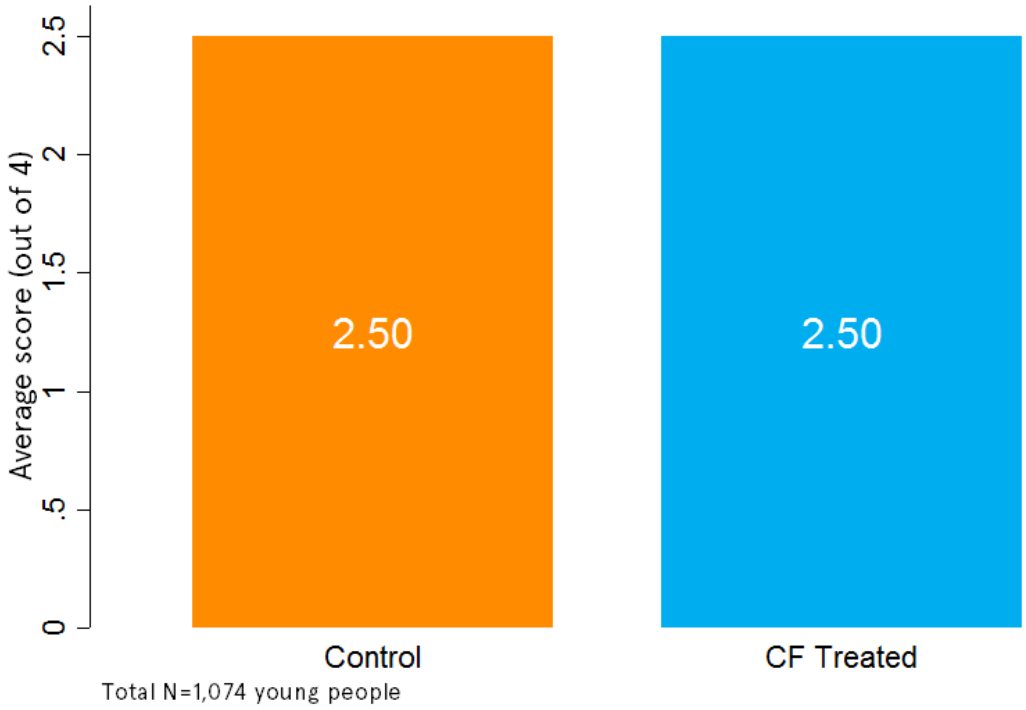


Figure 3: Citizenship Foundation social trust scores

Envision results

The Envision programme showed statistically significant improvement on each metric tested, with the smallest change at 6%. Students who had participated in social action displayed a sense of community that was 16% higher than that of their counterparts in the control group, whilst empathy and cooperation were boosted by 11% each.

Note that the post-survey was, necessarily, taken part way through the programme as the duration of the Envision intervention is longer than the average initiative within the Fund and so the control group schools needed to start their social action project before the treatment schools finished theirs. As a result, these findings were obtained using data collected before the project's completion.

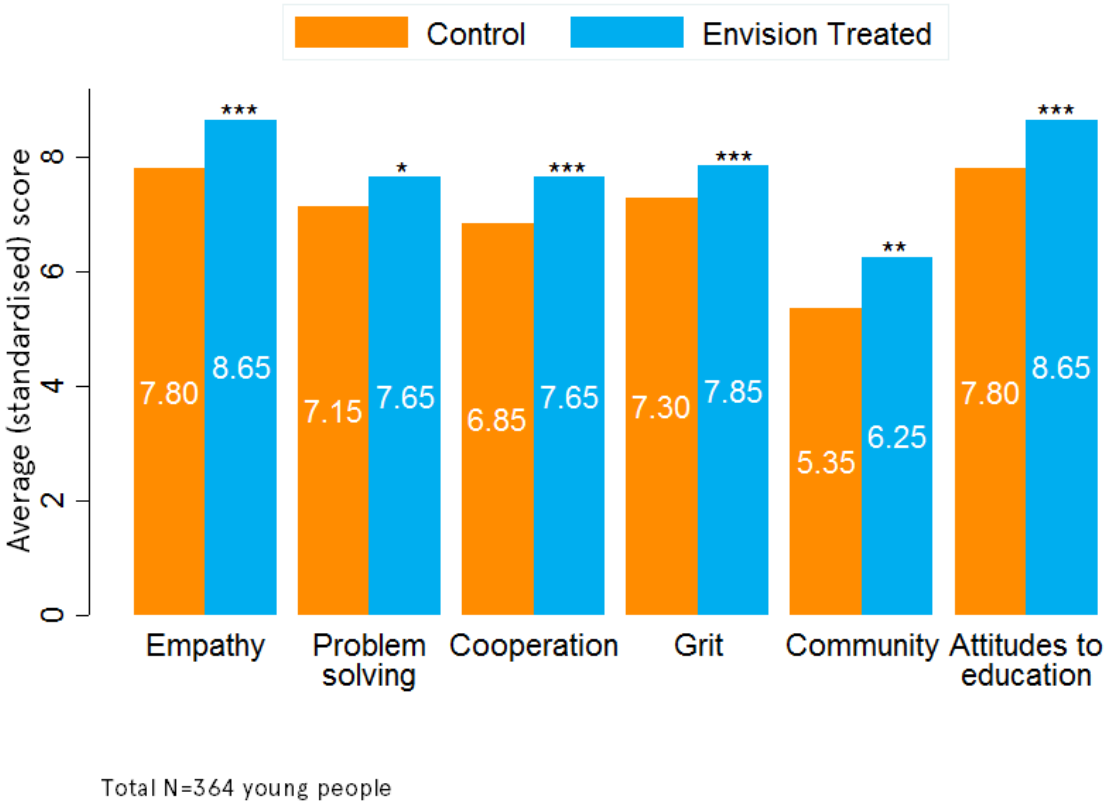


Figure 4: Envision results for the six psychological constructs

The results on wellbeing scores are also positive, showing that those who participated are more satisfied with life and see the things in its as being worthwhile. We did not survey the Envision participants on social trust.

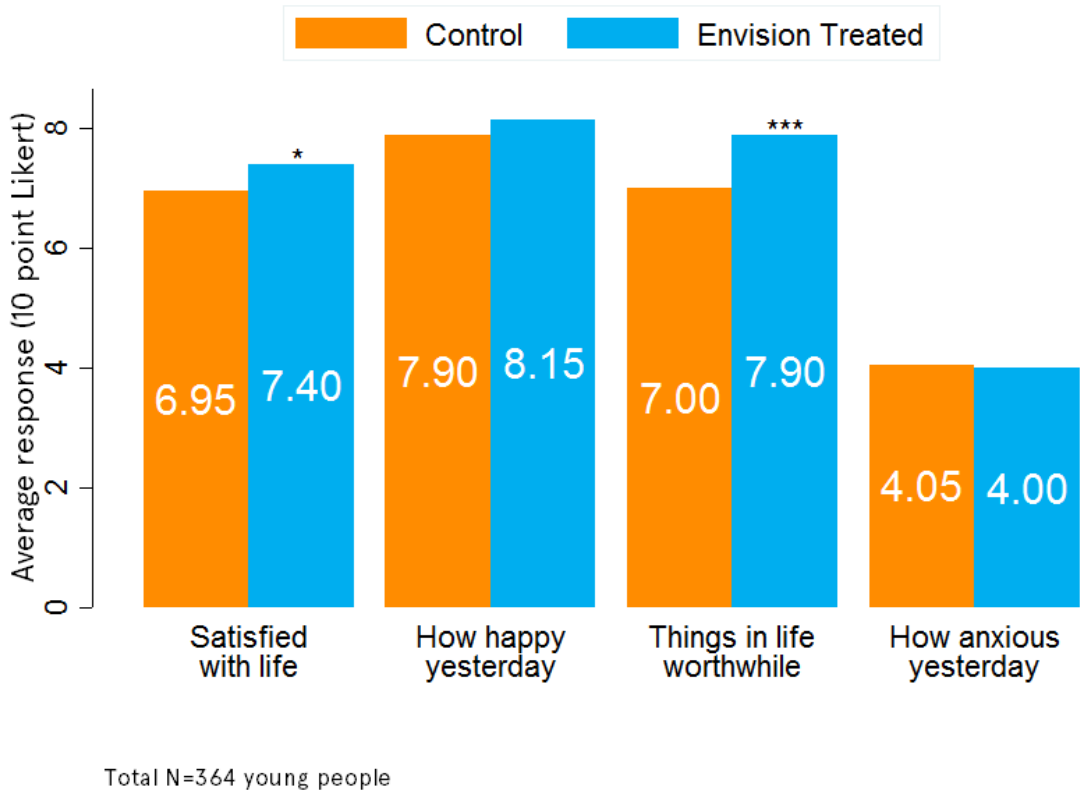


Figure 5: Envision results on wellbeing levels

Voluntary Action Within Kent (VAWK/IMAGO) results

VAWK/IMAGO's programme also realised significant improvements in empathy, cooperation and community involvement. Students who participated in VAWK/IMAGO's social action programme showed an 8% increase in levels of empathy and a 9% increase in cooperation compared to students who didn't participate. Similarly, those students who participated had considerably higher community involvement: 5.3 relative to the control students' 4.7, an increase of 15%.

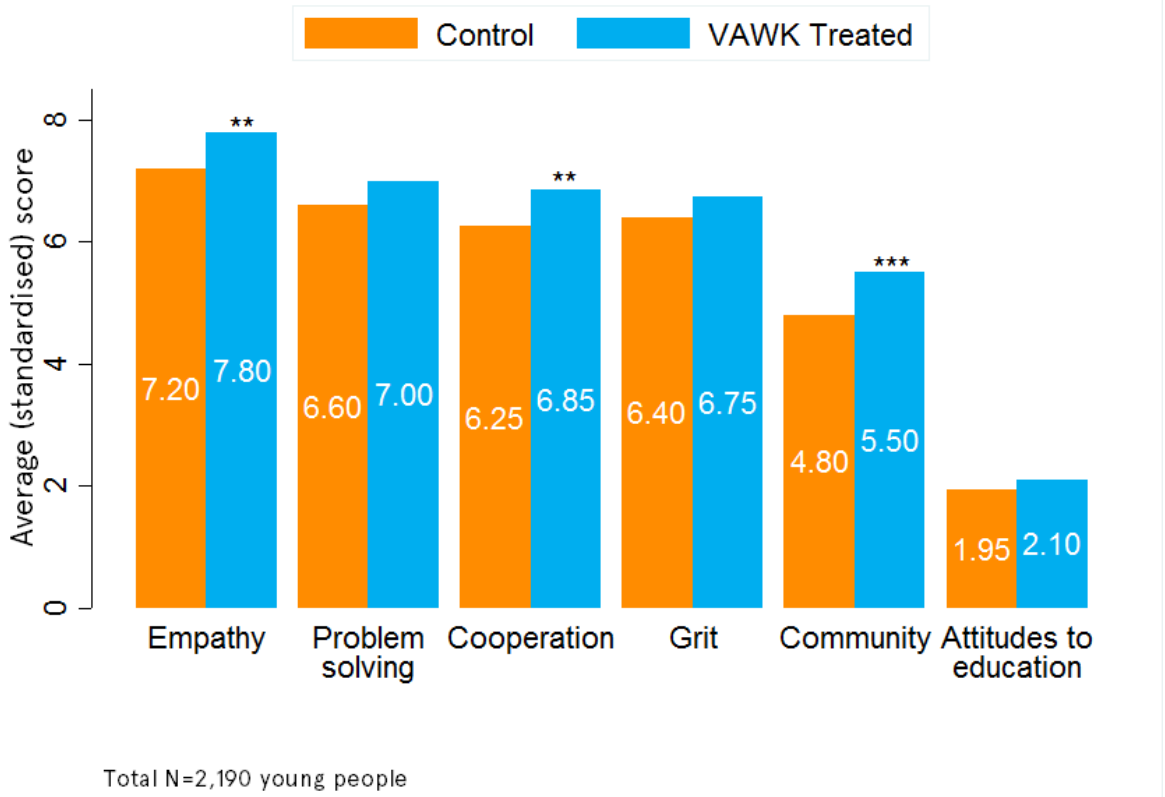


Figure 6: VAWK/IMAGO results for the six psychological constructs

As with the other programmes, VAWK/IMAGO participants reported increased wellbeing through higher levels of agreement with the sentiment that things in life are worthwhile. They also reported higher levels of happiness and lower levels of anxiety than their counterparts, although only the latter of these is statistically significant.

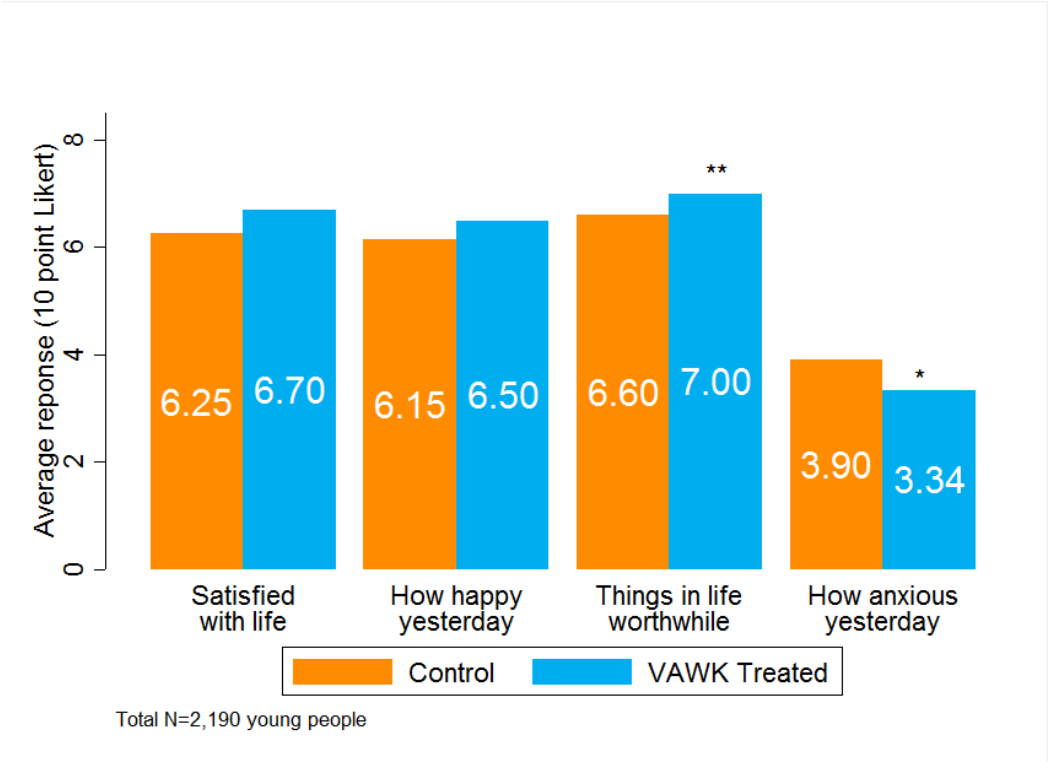


Figure 7: VAWK/IMAGO results on wellbeing levels

Finally, the VAWK/IMAGO participants did display a significant increase in social trust; a key ingredient for success in adult life.

Figure 8: VAWK/IMAGO social trust scores

Expressions of willingness to volunteer in the future

We measure whether young people who participate in social action are more willing to participate in volunteering thereafter. Additionally, we explore whether young people had a greater willingness to help those in their own community (an "in-group") as well as unrelated individuals (an "out-group"). Participants from both treatment and control schools were asked to indicate interest in further volunteering based on their (yes/ no) responses to two statements:

i. I would like to learn more about how I can spend my own time helping people in my community in the future.

ii. I would like to learn more about how I can spend my own time helping people in other countries in the future.

The former measures in-group willingness, the latter out-group willingness. We find that previous experience of social action – that is to say, taking part in one of our Provider initiatives – significantly boosts interest in volunteering opportunities again in the future, although it is clear that this is not the same as taking action. The graphs below show this effect for each programme individually.

Young people who participated in the Citizenship Foundation programme were significantly more willing to participate in in-group volunteering. Notably, however, there was no difference in willingness to volunteer with out-groups based on students' participation.

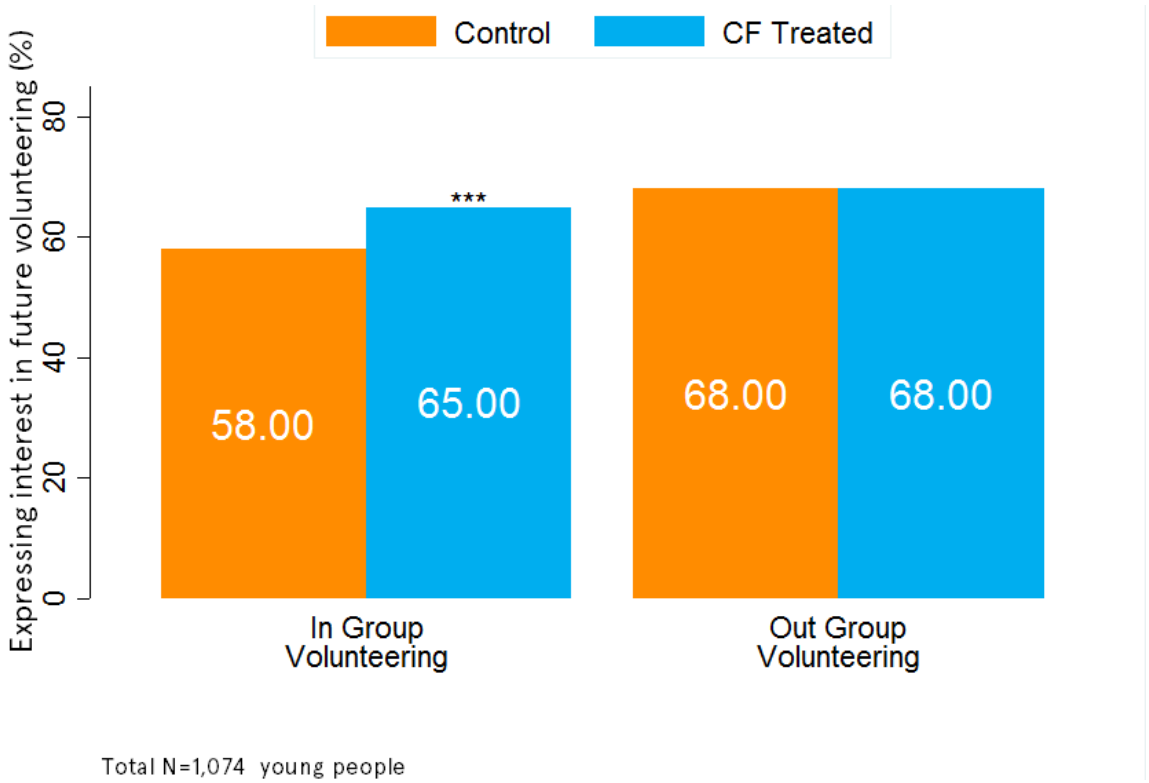


Figure 9: Citizenship Foundation willingness to volunteer in future results

As with the social trust question, one possible explanation for the similarity of responses for the option to volunteer abroad is age of the children taking part in the Citizenship Foundation initiative. The children are relatively young – ten year olds – and so may have fixed or unrealistic conceptions about what volunteering abroad would entail that are unlikely to move until they gain more experience.

Those who participated in Envision's social action were considerably more likely to be willing to volunteer with both in-groups and out-groups. Overall, those who participated in Envision's programme were more than 20 percentage points more likely to express a willingness to volunteer than those who did not participate.

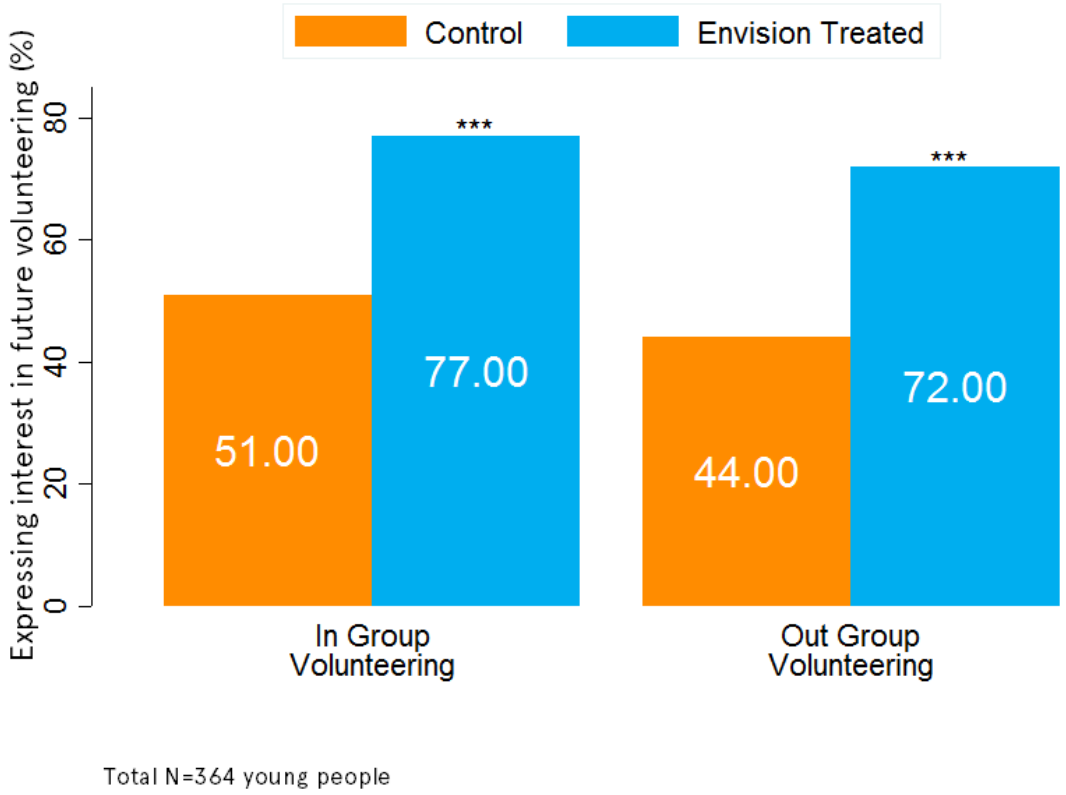


Figure 10: Envision willingness to volunteer in future results

Similarly, the young people who participated in VAWK/IMAGO's social action programme also indicated higher levels of willingness to volunteer in both in-group and out-group populations. Those who were in the treatment group were more than 10 percentage points more likely to express interest in the opportunity.[11]

[11] It is interesting, and not something to be explained in this report, that the baseline levels of willingness to volunteer amongst VAWK/IMAGO students (both treatment and control) are much lower than amongst the other Programmes. This suggests some kind of systematic difference in the students relative to the other two groups (which are, in turn, very different from one another). This in turn underscores the importance of trialling initiatives in various different locations.

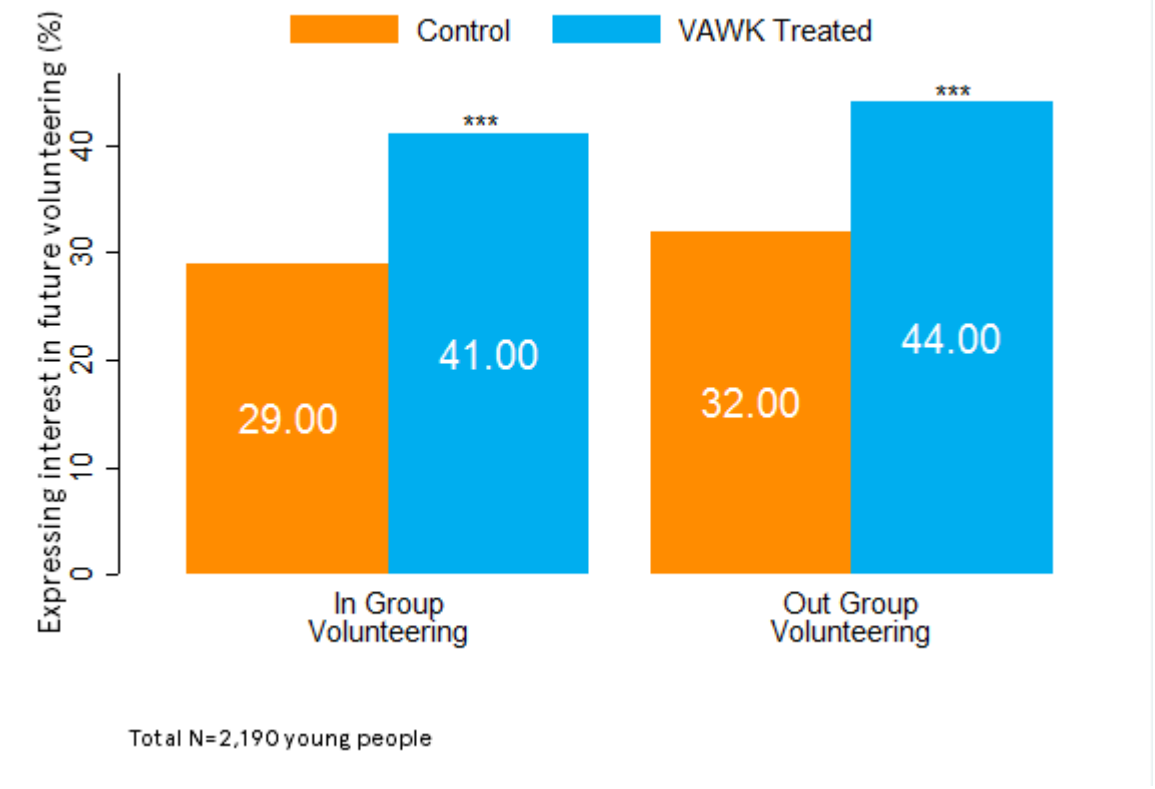


Figure 11: VAWK/IMAGO willingness to volunteer in future results

Behavioural trial

Willingness of Citizenship Foundation participants (treatment and control) to donate to charity

As part of the quest to discover whether taking part in social action leads to further pro-social behaviour in the future, we gave young people in the Citizenship Foundation treatment and control groups the opportunity to donate money to charity. We gave each child four 50 pence pieces and two envelopes; one marked "me" and one marked "charity". We then told the young people to put the money in whichever envelope they wanted and in whatever proportions they chose. They were then allowed to keep the "me" envelope.

Interestingly, whilst the effects of undertaking social action projects in the past strongly motivate a desire to volunteer in future (see above), the opposite was true of donations. We found that those who had taken part in the initiative donated significantly less money to charity than their peers who had not taken part in the initiative. Treatment students gave almost 16% less money than those in the control group.

It is possible that this is down to a licensing effect ("I've already done something good so I deserve this") but it could also be a result of a number of other factors

such as the children's conception of the relative value of giving time instead of money. The graph below shows these results.

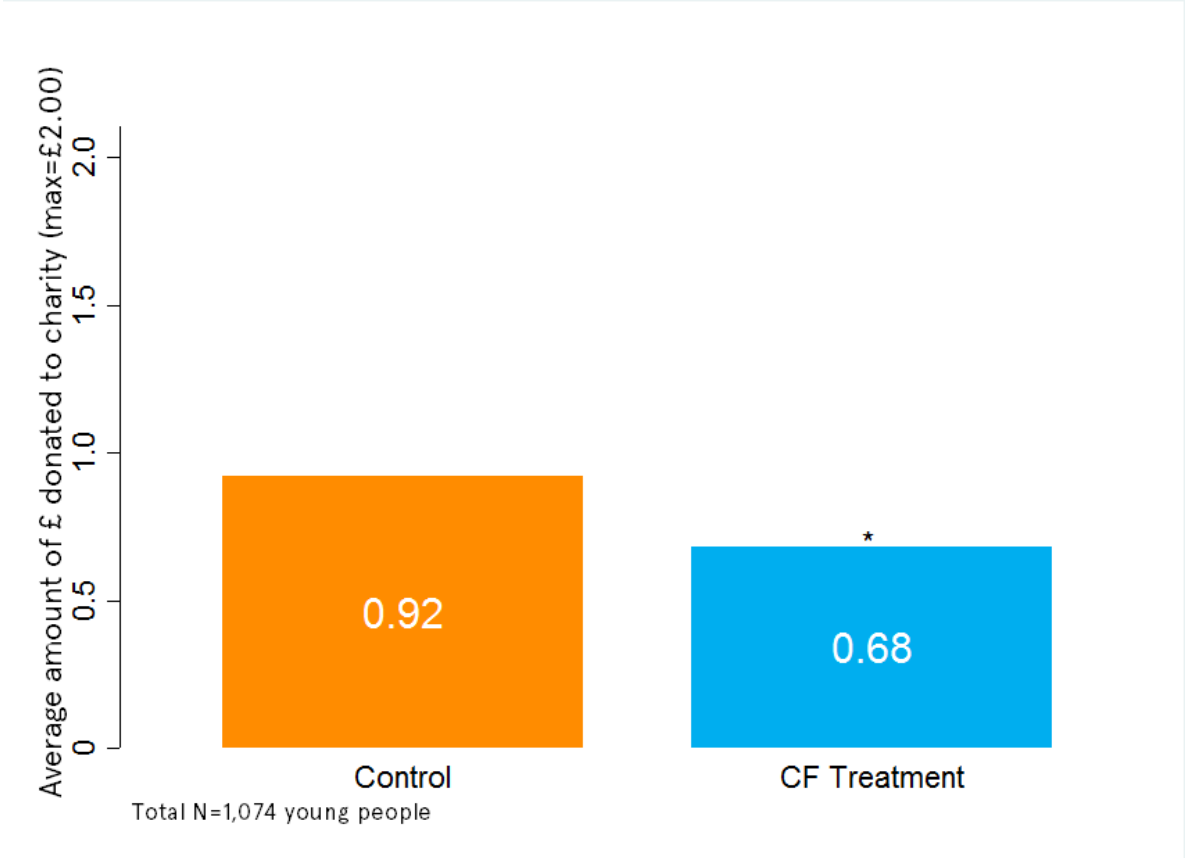


Figure 12: Average number of 50 pence pieces donated to charity per student

Interview Measure of Employability

One of the major goals of the evaluation was to find out whether taking part in social action increases students 'soft skills' that can translate into employability. To test this, we conducted a number of mock interviews with young people in the treatment and control groups of the Envision Trial.

These interviews form a variant of the Trier Social Stress test, which is routinely used in psychology to test soft skills and responses to moderately stressful environments, like an interview. The test was modified to make it more suitable for a young audience, specifically by having only one interviewer present, who did not take notes during the interview. Overall, 45 interviews were recorded – 27 in the control group and 18 in the treatment group. The sample came exclusively from the Envision group. The decision to do this with Envision was based on availability of a sample and appropriateness (CF participants were deemed too young to take part in a measure simulating a job interview), and the sample was selected based on the size of previous pieces of work using the Trier Social Stress test.

These short interviews were filmed, and subsequently coded on two separate occasions. In the first instance, it was done by 12 independent coders – a mix of British and American coders with experience conducting interviews and rating candidates. In the second instance, it was done by 65 independent coders – with more varied experience of conducting interviews, and who were asked to respond based on a general impression they had of the candidates. Coders were asked to decide whether they would hire the person in the interview, answering either "Yes", "No", or "Maybe" Each coder watched all interview segments, presented in a random order to control for decision fatigue (the tendency to rate either less attentively, or more negatively, as time goes on) and anchoring effects (whereby the first interview seen influences reviewers perceptions of subsequent interviews).

The results of pooling these ratings are shown in the figure below.

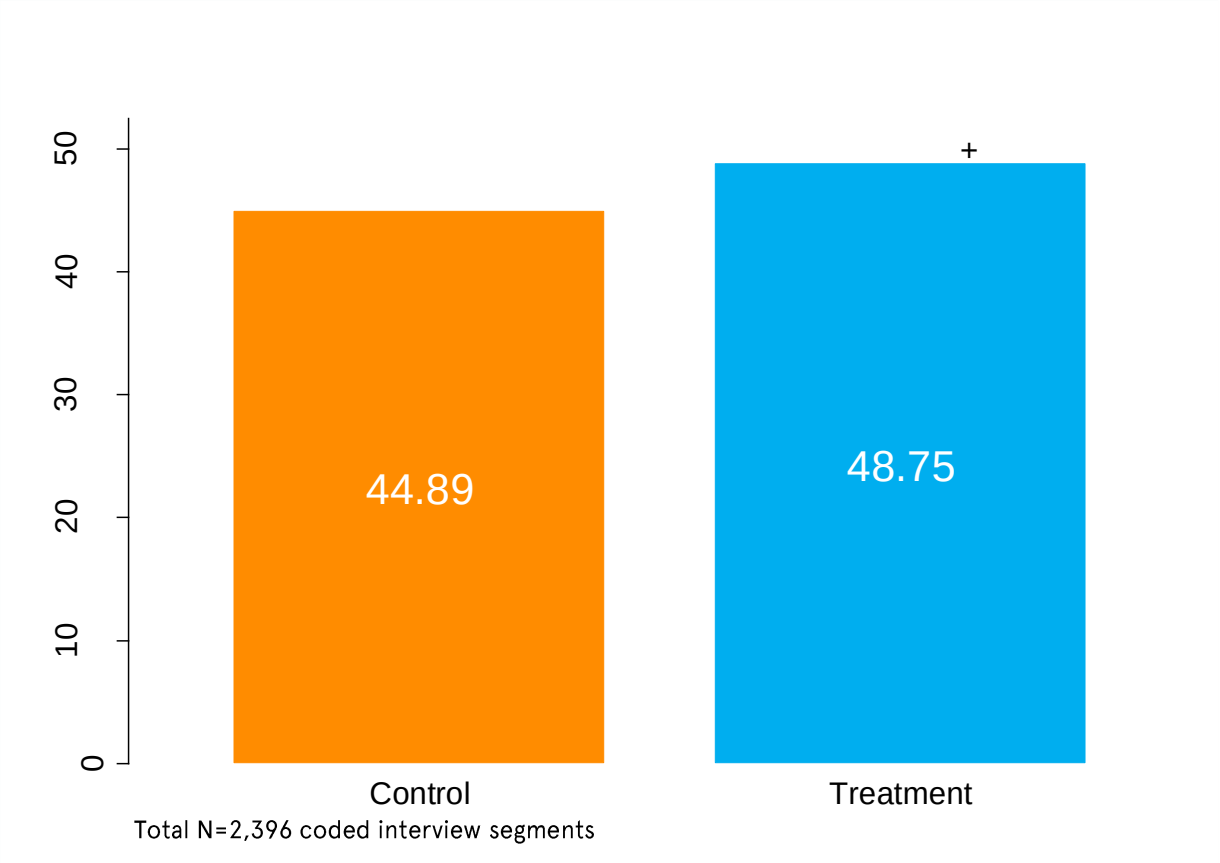


Figure 13: Results for the Interview Measure of Employability

As can be seen, we find some evidence that taking part in social action boosts employability according to the interview measure of employability. However, this measure is only statistically significant at a 0.10 alpha level (indicated by the cross

above the treatment bar), and so should be treated with some caution. However, results from the 12 experienced coders pointed in the same direction as the results from the 65 Mturk coders – and the significance is increased when pooling the results.

Conclusion

This report shows a number of promising results in relation to both social action and the measurement of its effect.

Primarily, this evaluation was crucial in providing robust evidence that, quite aside from its primary benefits, youth social action has a positive impact on building the skills its participants need for life and work. Though the programmes observed were each different in their approach to meeting Step Up To Serve's six principles of social action, they consistently improved young people's levels of empathy, and community involvement. Particular programmes were also impressive in increasing students' cooperation and levels of grit. Moreover, those who participated in social action seemed largely more willing to donate their time in the future, though no such effect was seen with money.

These results are important in indicating that investment in youth social action may have lasting effects on engagement with other forms of social action. These positive results confirm that the Step Up To Serve #iwill campaign has an important role in encouraging young people to take part in social action within their communities, and in influencing organisations to offer more opportunities. Furthermore, the returns from youth social action can be quantified; welcome news to social investors looking to fund initiatives that make a concrete difference.

Additionally, this evaluation paves the way for others in the future by providing helpful tools. The surveys used to evaluate the six traits and characteristics in which we are interested are now validated except one metric, students' educational attitudes. As such we can be certain that the remaining questions measure what we intend to measure and do so reliably. This provides a basis on which to evaluate future social action initiatives.

While these results are promising, it is important to continue collecting data so that the longer-term outcomes of youth social action are known. In particular, observing National Pupil Database outcomes on attendance and attainment down the line would provide insight on the degree to which social action impacts educational attainment even years after the programme has been completed. Moreover, similar evaluations of other related initiatives would be helpful in testing the quality of those programmes as well as understanding other important outcomes of interest and how applicable these results are in other settings and with different participants.

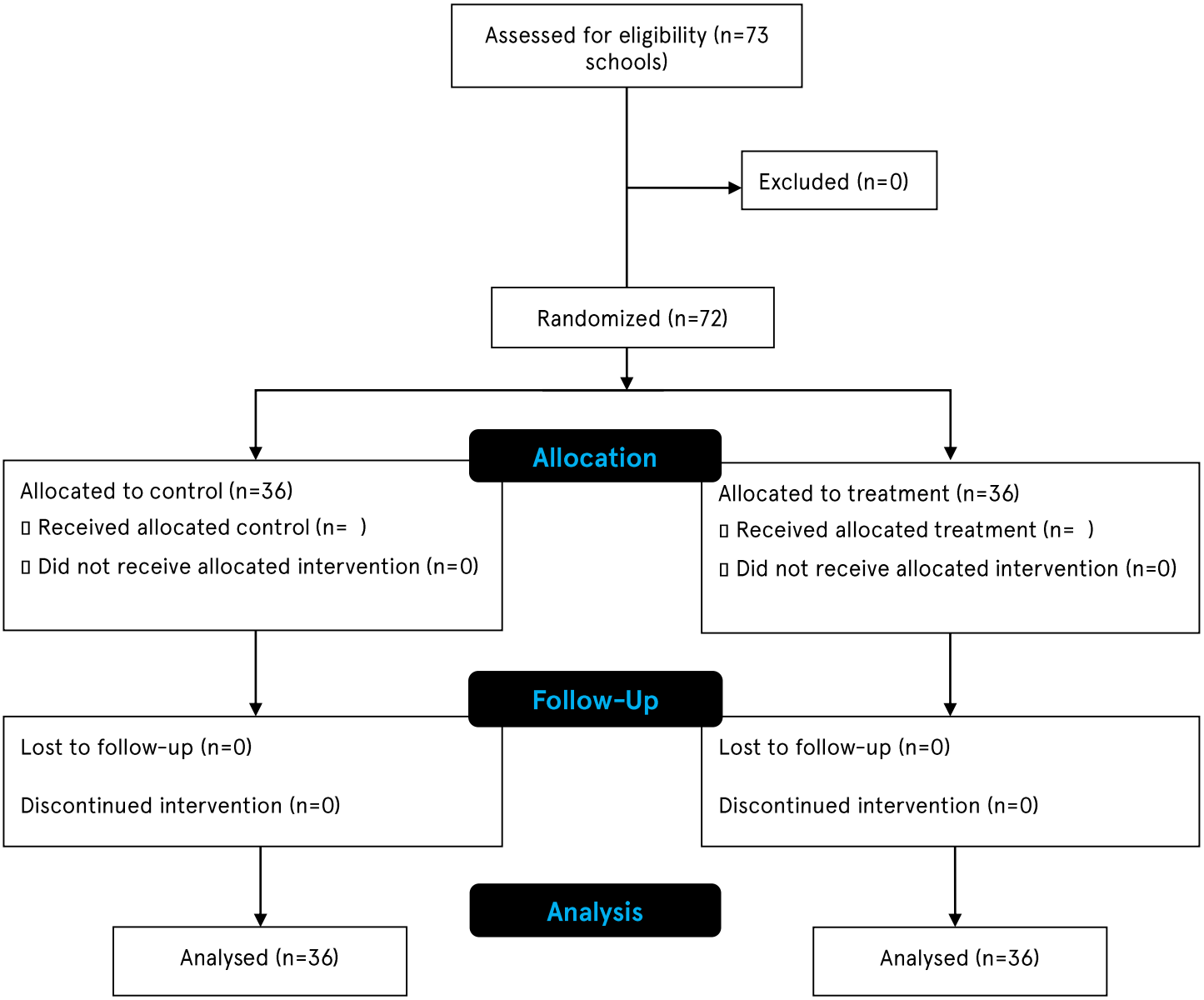
Technical Appendix

Randomisation Method

So that one students' experience would not alter that of a peer in the same school, randomisation was performed on the school level. For all trials, an entire school was either offered the social action programme or an entire school was not. Randomisation was stratified on the school's 'type' – for example whether or not they were an academy, and whether or not the school was religious. In those schools that were selected to be treatment schools, students were offered an opportunity to participate in the social action programme, but were not obliged to. The sample of control students was composed of students in a classroom randomly selected by the school. Note, though that only those students who express interest in participating in a social action program are included in the control sample so as to ensure that the there is no selection bias in the treatment group. It is recognised that this presents external validity issues for these trials; both the schools and the students opted to take part (although not whether they would be in the treatment or control group). As such, findings indicate the results of these specific programmes in schools that value social action and amongst students that opt into these activities.

In order to ensure that the characteristics of the schools were balanced between the treatment and control groups, randomisation was stratified. That is, schools were grouped by religious affiliation and school type and then randomised within those groups. This randomisation occurred using a random 0/1 number generator.

CONSORT flow diagram of attrition rates (where unit of randomisation = participating schools)



Power Calculations

Power calculations were conducted for each programme in advance of their launching in order to establish whether the randomised trials had enough statistical power to detect plausible changes in the outcome measures.

These calculations are helpful, in that they allow us to determine the likelihood of a trial succeeding, and hence whether or not the programme should be evaluated in this way. For the three RCT programmes, power was calculated on the basis of attempting to address a continuous outcome measure (an approximation to a 30 or 40 unit categorical variable), in which randomisation, and hence outcomes, are clustered at the level of the school. Clustering, although the only practical method for randomising in the case of this trial, reduces statistical significance because levels of the outcome measure are naturally correlated within a school, and so this must be accounted for by corrections to the trial's standard errors. Our power calculations account for some level of intra-cluster-correlation, which is specified at 0.02 for these trials, consistent with the literature in this area[12].

These calculations produce the figures found in the table below, which are expressed in Cohen's D (following Cohen, 1988), which expresses effects in terms of standard deviations of the outcome measure. Cohen defines a small effect size as being between 0.2 and 0.3 standard deviations, a medium effect size as being 0.5 standard deviations, and a large effect size as being 0.8 standard deviations. By these standards, the Citizenship Foundation and VAWK/IMAGO RCTs are powered to detect small effects, and Envision is powered to detect effects that are between small and medium in size.

Power Calculations				
Programme	N	Number of Schools	Design Effect	Detectable effect size
Citizenship Foundation	1074	31	1.672	0.22
Envision	364	16	1.436	0.35
VAWK	2190	25	2.732	0.19

[12] Kerry, S. M., & Bland, J. M. (1998). The intracluster correlation coefficient in cluster randomisation. *BMJ*, *316*(7142), 1455-1460.

Selecting measures

In selecting measures, we had several broad challenges to address. First, we needed questions that reliably measured employability in a population not seeking to work. Second, we needed to evaluate the processes as well as the concrete outcomes to capture the full breadth of value. And third, we hoped to do both of the above without placing a heavy burden on participating schools and survey respondents.

To identify which outcomes are suitable indicators of employability, we started with the Quality Outcomes Framework and supplemented this with research on what precedent exists in the literature and research, the practical aspects of what providers could and could not measure, and qualities that could be measured without having to ask young people too many question

After choosing outcomes, we went through a similar process to design measurement scales based on:

1. Validity of measure
2. Size
3. Use in other programmes so that we can benchmark YSA against other existing policy programmes
4. Consistency with the EEF programmes
5. Capacity to detect change

To ensure that selected questions are age specific and that scales use a consistent measurement throughout the survey, we adapted some of the questions used in previous studies.

Overall outcome	Outcome	Coverage in evaluation	Scale/source
Employability	Y/N mock-hire decision; overall survey responses	Interviews and surveys	Interview assessments based on hiring coding sheets and, separately, the view of those experienced in hiring young people. All survey measures.
Pro-social behaviour	Post-programme interest in YSA/volunteering and donations to charity	Post-programme follow-up soliciting interest in future volunteering and donations game	Expressions of interest in future volunteering; charitable donations made when given the means and opportunity
Education measures	Attainment	Attainment (NPD)	National Pupil Database-linked using Unique Pupil Number (UPN)
	Behaviour	Attendance (NPD)	
		Exclusions (NPD)	
	Communication	Pre/post survey- Q 3.04,3.06	California Healthy Kids Survey. Resilience and Youth Development Module- Internal Assets
	Problem solving	Pre/post survey- Q 3.02, 3.04	
Personality Traits	Cooperation	Pre/post survey- Q 5.01-5.03	
	Grit	Pre/post survey- Q 5.04-5.06	
	Empathy	Pre/post survey- Q 5.07-5.09	
	Problem Solving	Pre/post survey- Q 5.11-5.12	
	Grit	Pre/post survey- Q 5.13	Locus of control- National Citizenship Service

Community Involvement	Previous involvement	Pre/post survey- Q 7	Community Involvement- Longitudinal Study of Young People in England (LSYPE), Volunteering Module
	Frequency of involvement	Pre/post survey- Q 8	
	Community impact	Pre/post survey- Q 9.01- 9.04	Community Measures- National Citizenship Service
	Community cohesion	Pre/post survey- Q 9.05	
Social Capital	Trust	Pre/post survey- Q 10	Community Life Survey 2012-2013- Your Community module

Instrument

The table outlines the questions we used to identify each of the 6 constructs of interest. The questions were asked in exactly this way, regardless of the young person's treatment group.

Construct	Question (agreement with statements is rated on a scale)
Empathy	I feel bad when somebody gets their feelings hurt
	I try to understand what other people go through
Problem Solving	I know where to go for help with a problem
	I am confident about having a go at things that are new to me
Cooperation	I can work with someone who has different opinions to me
	I enjoy working together with other students my age
Grit and resilience	I often figure out different ways of doing things
	If something goes wrong I am able to bounce back and carry on
	Once I have started a task, I like to finish it
Sense of Community	I feel able to have an impact on the world around me
	I feel motivated to take action on issues in my community
Attitudes to Education	I'm not interested in doing any more learning
	Studying to gain qualifications is important to me

Descriptive statistics

Citizenship Foundation			
	Treatment allocation		
Gender	Control	Treatment	Total
Male	300 (47.32%)	334 (52.68%)	634 (100%)
Female	299 (48.94%)	312 (51.06%)	611 (100%)
Total	599 (48.11%)	646 (51.89%)	1,245 (100%)

Envision			
	Treatment allocation		
Gender	Control	Treatment	Total
Male	85 (57.05%)	64 (42.95%)	149 (100%)
Female	96 (44.44%)	120 (55.56%)	216 (100%)
Total	181 (49.59%)	184 (50.41%)	365 (100%)

Citizenship Foundation			
Age	Treatment allocation		Total
	Control	Treatment	
8	1 (100%)	0 (0%)	1 (100%)
9	2 (100%)	0 (0%)	2 (100%)
10	87 (49.71%)	88 (50.29%)	175 (100%)
11	268 (42.34%)	365 (57.66%)	633 (100%)
12	222 (55.09%)	181 (44.91%)	403 (100%)
13	2 (100%)	0 (0%)	2 (100%)

Envision			
Age	Treatment allocation		Total
	Control	Treatment	
16	0 (0%)	1 (100%)	1 (100%)
17	31 (46.97%)	35 (53.03%)	66 (100%)
18	135 (50.37%)	133 (49.63%)	268 (100%)
19	13 (50%)	13 (50%)	26 (100%)
20	0 (0%)	1 (100%)	1 (100%)

Instrument Validation

In order to ascertain whether the questionnaire was assessing the metrics intended, we sought to validate the instrument.

We first measured convergent by inspecting within-class correlation—namely that several questions measuring the same trait correlated with one another. To test the latter, we inspected between-class correlation for those constructs the literature suggests are less related. Those were: (1) empathy and grit/ resilience, (2) problem solving and sense of community, and (3) cooperation and attitudes toward education. Given the uncontrolled context of the study, more lengthy attempts at validation would be unlikely to yield meaningful results and may risk spurious precision. Correlations were analysed following Campbell & Fiske (1959) to determine if the two sets were measuring theoretically different constructs.

We began by testing the reliability of each measure. We did so by calculating Cronbach's α for each of these measures – these are displayed below.

Cronbach's Alpha			
Measure	Envision Control	VAWK Control	Aggregate
Empathy	82	77	78.1
Problem Solving	55	52	52.7
Cooperation	57	62	60.9
Grit and resilience	86	72	75.1
Community	75	59	62.5
Attitudes to Education	41	34	35.5

As a rule of thumb, scores of less than 50 are deemed unacceptable, while scores of between 50 and 60 are deemed poor. Scores above this level are either acceptable or good. By this measure, our questions to identify attitudes to education appear do not appear to be reliable, and the reliability of the problem solving questions is also relatively poor. We note that downward bias in Cronbach's α can occur due to a small number of questions (which is certainly an issue here), and because of ceiling or floor effects, which are particularly an issue in our sample for measuring attitudes to education (the majority of participants score 1/10 when answering a question on education). Although we therefore have cause

to question the reliability of attitudes to education questions, we do not find much to be concerned about in the reliability of the other measures. We do note that in future uses of this instrument, additional measures should be included to more accurately gauge educational attitudes.

We also tested divergent validity; checking that those constructs which should not relate to one another do not, indeed, correlate. Discriminant Validity is calculated based on the correlation between items within a scale, and comparing these to the correlation between items of the comparison scale. Hence, for each of our constructs we have a score, calculated:

$$\frac{Correlation\ Between\ Construct\ and\ Comparison}{\sqrt{(Correlation\ within\ Construct \cdot Correlation\ within\ Comparison)}}$$

In this test the pairs of constructs used are; empathy and grit, problem solving and community, & cooperation and attitudes to education. The graph below shows the score calculated for each of the constructs, where the construct labelled shows the primary point of comparison. Scores within pair vary as a result of the different numbers of items within the various scales.

Scores less than 0.85 are considered to be sufficiently divergent. However, our two weakest measures in the reliability test, Problem Solving and Attitudes to Education, come the closest to failing under this test as well.

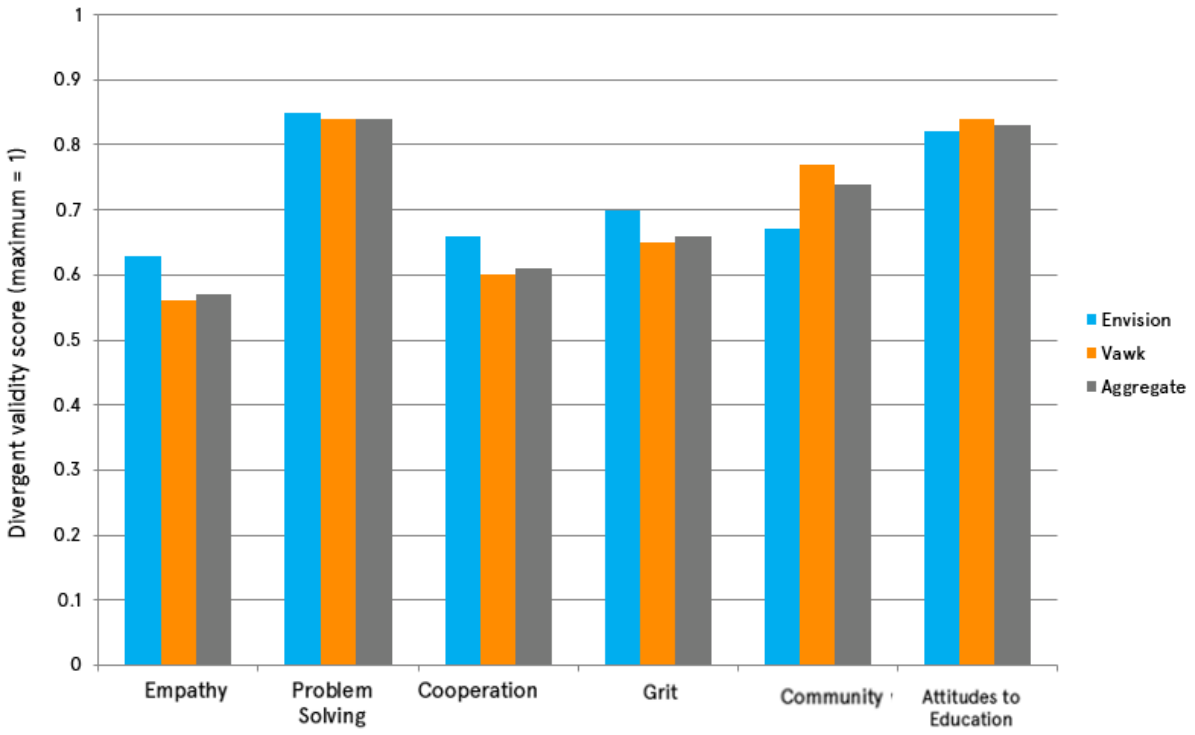


Figure 94: Divergent validity scores by construct by trial

Analytic Strategy – Randomised Trials[13]

For self-report measures we estimate an ordinary least squares (OLS) regression model:

$$S_{ic} = \alpha + \beta_1 T_{ic} + \beta_2 C_c + \beta_3 X_{ic} + u_{ic}$$

Where S_{ic} represents our outcome measure (such as empathy), measured using a self-rating scale that ranges from 0 to 10 on a Likert scale,

α is a constant,

T_{ic} is a binary variable set to 1 if an individual is in the experimental period and 0 otherwise ($C \in \{0,1\}$). Hence, β_1 can be interpreted as the difference between the control and treatment group (the treatment effect on the treated),

C_{ic} is a vector of school characteristics,

X_{ic} is a vector of participants characteristics, and

u_{ic} is an i.i.d. error term, clustered at the school level.

Analysis of CSV programme

Data for the CSV trial are limited. In total, only 22 participants' post-surveys were provided, all from the same school. All of these were paired with the same student's response from the pre-survey. The possibility of matching participants from this school with participants with other schools based on their characteristics and pre-survey responses was considered and investigated. However, due to the relatively uncommon characteristics of the school among the remainder of our sample, (a girls-only selective school in Kent), and the generally high levels of all of the aggregates in this school's pre-period data (typically higher than the post treatment levels for this same variables in other schools), matching was not considered to offer a valuable counterfactual, potentially increasing bias, but could risk artificially enhancing the precision of estimated treatment effects. As noted above, all participants in this school are female, and so we estimate a simple two period model;

[13] The samples for CF and Envision are too small for sub-group analysis. With the standard errors being clustered, the detectable effect size become too large. With the VAWK data, most of the effects are driven by the girls, but for most outcomes the difference between boys and girls treatment effects is insignificant. There are also no significant interactions between treatment and age.

$$S_{it} = \alpha + \beta_1 T_{it} + u_i$$

Where S_{it} represents our outcome measure (such as empathy), measured using a self-rating scale that ranges from 0 to 10 on a Likert scale, For each individual the subscript t can take two values, 0 or 1, depending on whether the participant is being measured in the pre or post periods.

α is a constant,

T_{it} is a binary variable set to 1 if an individual is in the experimental period and 0 otherwise ($C \in \{0,1\}$). Hence, β_1 can be interpreted as the difference between pre period and the post period – the change in the outcome measure, and

u_i is an i.i.d. error term, clustered at the level of the individual participant (for whom there are two observations).

Community Service Volunteers results:

CSV's programme did not produce significant improvements across any of the aggregated measures captured by our survey (see Figure 15). Although the majority of measures show movement in the desired direction, the small sample size available (22 pre and 22 post survey observations), makes the measurements too imprecise to achieve statistical significance. As a result, neither the positive results, nor the negative estimated effects for problem-solving and co-operation should be viewed too seriously.

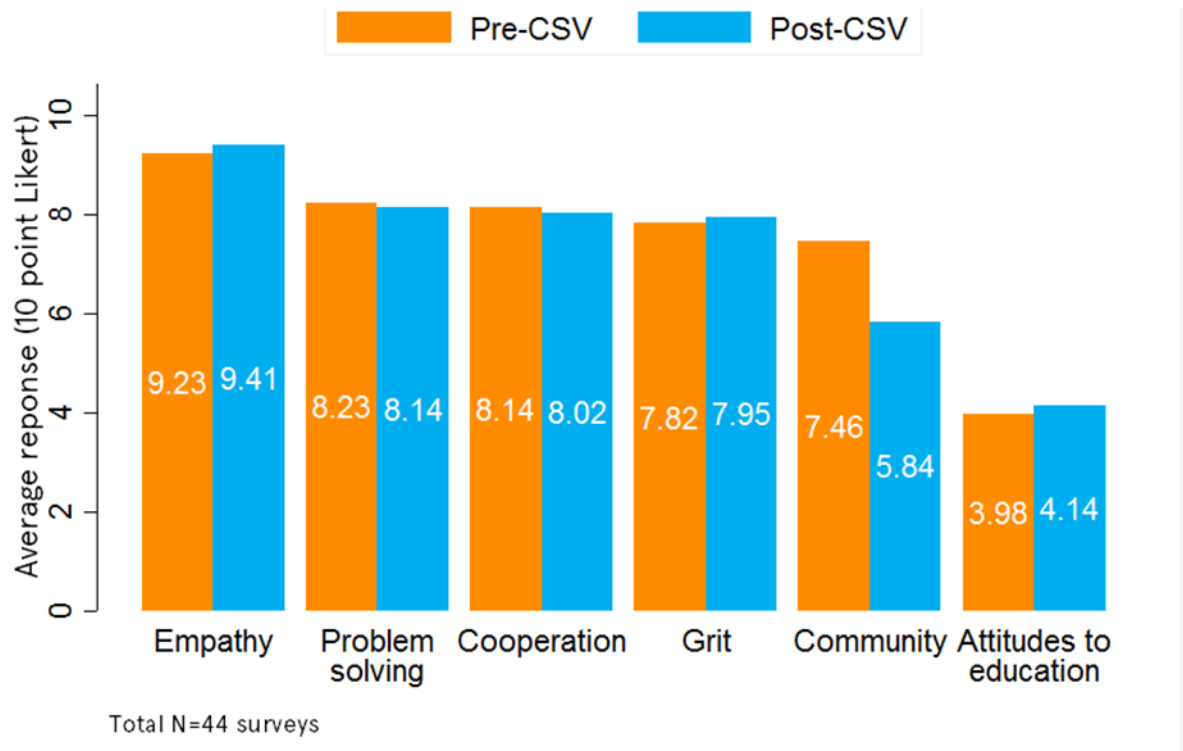


Figure 15: CSV results for the six psychological constructs

As with the other programmes, CSV participation is associated with higher levels of life satisfaction and happiness. Participants views that life is worthwhile did not change over the period of the programme, a while anxiety fell only very slightly.

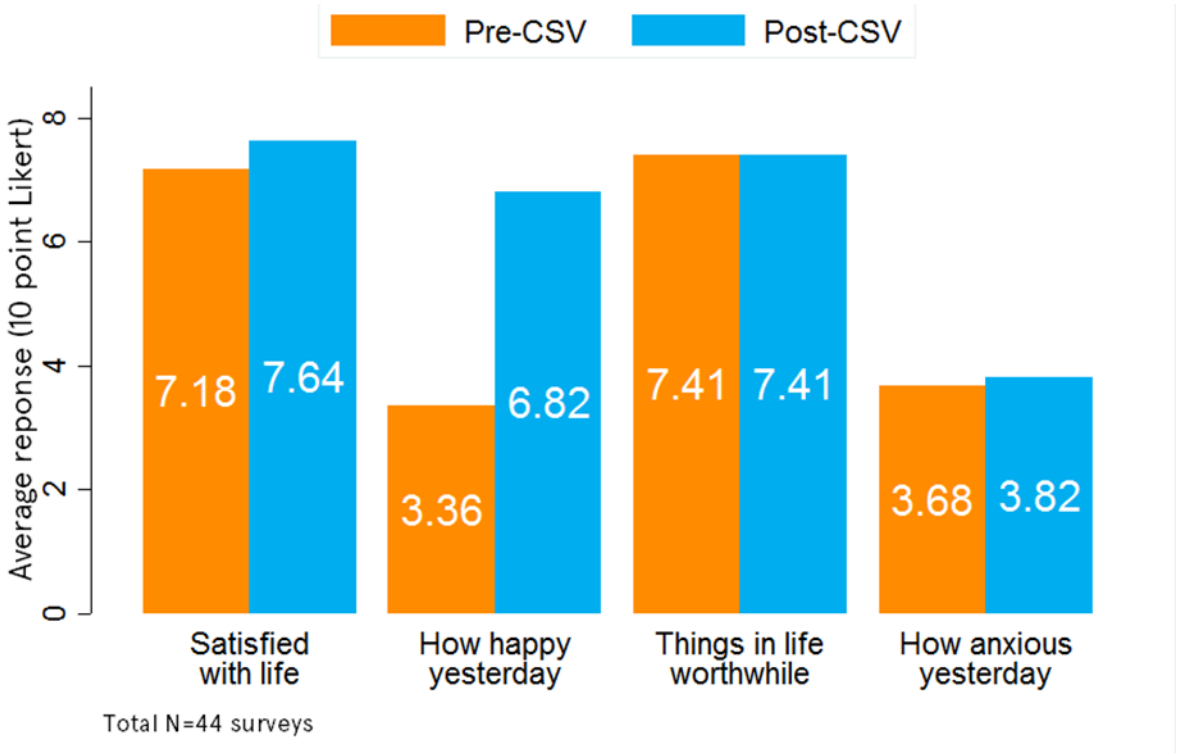


Figure 16: CSV results on wellbeing levels

Finally, CSV participants display a small and statistically insignificant fall in their levels of social trust.

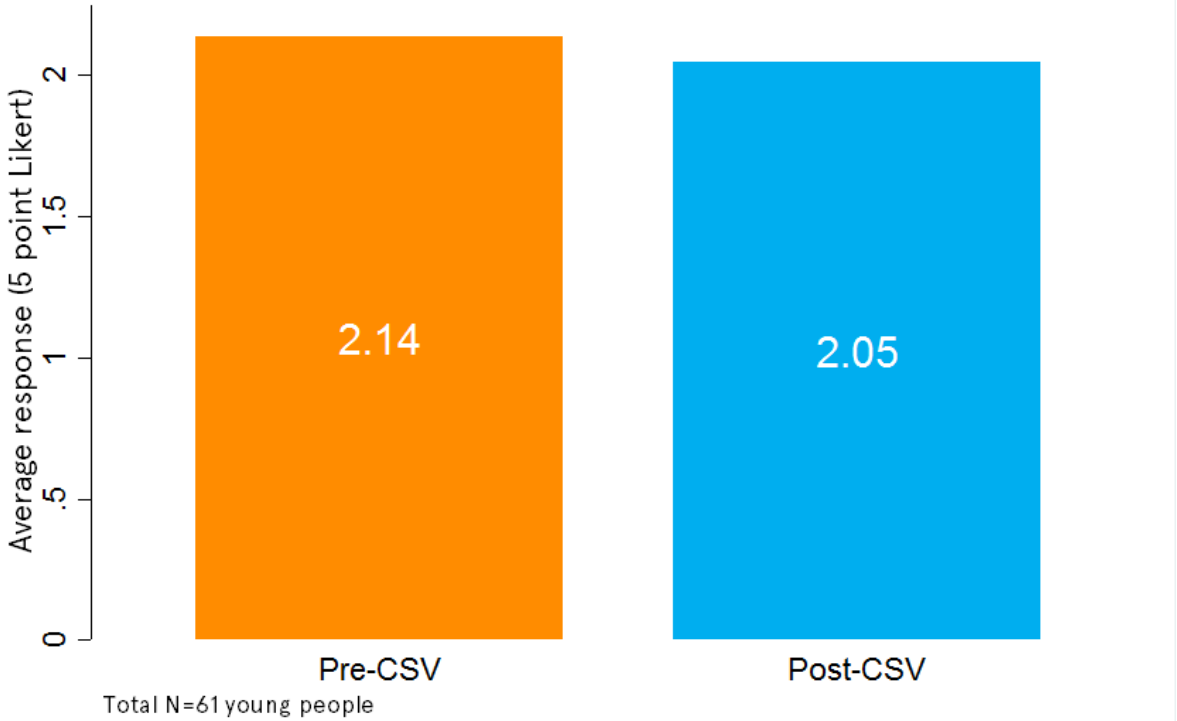


Figure 17: CSV Results on Social Trust levels

Analysis of Interview Measures:

The interview measure of employability was conducted using participants in the Envision programme. These participants were filmed answering a short series of questions in a mock interview.

A rating scale for the interviews was developed with two independent coders engaging in test-retest coding of randomly drawn subsets of the main sample (which also included similar interviews captured for other research), in concordance with a set of instructions. These instructions were modified until they created statistically similar results among both coders independently. Subsequent to this, these same instructions were used by 12 independent coders (not including the initial coders), to code all of the videos, on a 0,1,2, scale, where 0 indicated that the coder would "definitely not" employ this person, 1 indicated that they "may" employ this person, and 1 indicating that would "definitely" employ this person.

Our analytical strategy for this component of the evaluation is to estimate a linear prediction model:

$$H_{ic} = \alpha + \beta_1 T_i + \beta_2 \boldsymbol{f_c} + u_i$$

Where

H_{ic} is a binary outcome measure set to 1 if coder c scores participant i's interview as a 2 (would definitely hire) and 0 else,

α is a constant,

T_i is a treatment variable set to 1 if participant i is in the treatment group and 0 else,

$\boldsymbol{f_c}$ is a vector of coder fixed effects, and

u_{ic} is a standard error clustered at the participant level.

Results tables

Regression tables for each intervention are included below. Note that here as well, the outcomes are standardized to a 10-point scale.

Effects of CF Participation on Constructs

	(1)	(2)	(3)	(4)	(5)	(6)
	Empathy	Problem Solving	Cooperation	Grit	Community	Attitudes to Education
CF participation	0.427***	0.292**	0.056	0.579***	0.524***	-0.053
	(0.122)	(0.106)	(0.132)	(0.163)	(0.147)	(0.114)
Male	-0.511***	-0.091	-0.110	-0.205	-0.317*	0.280*
	(0.120)	(0.103)	(0.131)	(0.159)	(0.142)	(0.113)
Year 5	-0.011	0.167	0.208	0.033	0.544***	0.097
	(0.122)	(0.105)	(0.132)	(0.162)	(0.146)	(0.113)
Constant	7.059***	8.088***	6.701***	6.403***	6.291***	5.604***
	(0.193)	(0.176)	(0.217)	(0.269)	(0.228)	(0.186)
Observations	1074	1071	1074	1066	1059	1058

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

CF Wellbeing and Social Trust Qs

	(1) Satisfied with Life	(2) How happy yesterday	(3) Things in life worthwhile	(4) How anxious yesterday	(5) People can be trusted
CF participation	0.113	0.052	0.307*	-0.879***	0.003
	(0.153)	(0.187)	(0.145)	(0.225)	(0.044)
Male	-0.035	-0.161	-0.179	-0.523*	0.008
	(0.149)	(0.186)	(0.144)	(0.225)	(0.043)
Constant	7.737***	6.968***	7.311***	4.067***	2.325***
	(0.230)	(0.283)	(0.234)	(0.270)	(0.073)
Observations	1072	1074	1070	1067	1070

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Effects of Envision Treatment on Constructs

	(1) Empathy	(2) Problem Solving	(3) Cooperation	(4) Grit	(5) Community	(6) Attitudes to Education
Envision participation	0.852***	0.435*	0.766***	0.538**	0.859***	0.852***
	(0.203)	(0.199)	(0.182)	(0.187)	(0.233)	(0.203)
Male	-1.175***	0.202	-0.153	-0.394*	-0.047	-1.175***
	(0.229)	(0.206)	(0.187)	(0.200)	(0.236)	(0.229)
Constant	7.736***	7.165***	6.832***	7.317***	5.384***	7.736***
	(0.412)	(0.392)	(0.326)	(0.366)	(0.345)	(0.412)
Observations	364	362	364	358	353	364

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Envision Wellbeing and Social Trust Qs

	(1) Satisfied with Life	(2) How happy yesterday	(3) Things in life worthwhile	(4) How anxious yesterday
ENV participation	0.486*	0.229	0.896***	-0.061
	(0.220)	(0.271)	(0.216)	(0.334)
Male	0.227	0.168	-0.043	0.065
	(0.224)	(0.268)	(0.220)	(0.343)
Constant	6.949***	7.890***	7.030***	4.079***
	(0.351)	(0.411)	(0.322)	(0.506)
Observations	361	362	352	359

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Effects of VAWK Treatments on Constructs

	(1) Empathy	(2) Problem Solving	(3) Cooperation	(4) Grit	(5) Community	(6) Attitudes to Education
VAWK participation	0.560**	0.406*	0.543**	0.300	0.691***	0.213
	(0.176)	(0.180)	(0.166)	(0.175)	(0.166)	(0.245)
Male	-1.030***	0.429**	0.124	-0.142	0.307*	0.136
	(0.148)	(0.118)	(0.102)	(0.121)	(0.109)	(0.146)
Constant	7.284***	6.644***	6.364***	6.467***	4.658***	2.031***
	(0.168)	(0.116)	(0.140)	(0.132)	(0.198)	(0.269)
Observations	2190	2177	2189	2144	2157	2154

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

VAWK Wellbeing and Social Trust Qs

	(1) Satisfied with Life	(2) How happy yesterday	(3) Things in life worthwhile	(4) How anxious yesterday	(5) People can be trusted
VAWK participation	0.236	0.372	0.479**	-0.577*	0.182**
	(0.165)	(0.239)	(0.137)	(0.232)	(0.054)
Male	0.675***	0.803***	0.636***	-0.504**	-0.120**
	(0.110)	(0.134)	(0.105)	(0.167)	(0.036)
Constant	6.427***	6.097***	6.517***	3.876***	2.509***
	(0.177)	(0.240)	(0.143)	(0.132)	(0.067)
Observations	2175	2173	2155	2164	2144

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Effects of CSV Treatments on Constructs

	(1) Empathy	(2) Problem Solving	(3) Cooperation	(4) Grit	(5) Community	(6) Attitudes to Education
Change	0.182 (0.151)	-0.091 (0.190)	-0.091 (0.312)	0.136 (0.230)	0.386 (0.416)	0.159 (0.198)
Pre Score	9.227*** (0.193)	8.227*** (0.395)	8.114*** (0.388)	7.818*** (0.364)	5.455*** (0.453)	3.977*** (0.351)
Observations	44	44	44	44	44	44

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

CSV Wellbeing and Social Trust Qs

	(1) Satisfied with Life	(2) How happy yesterday	(3) Things in life worthwhile	(4) How anxious yesterday	(5) People can be trusted
Change	0.455 (0.709)	0.455 (0.878)	0.000 (0.731)	-0.045 (1.036)	-0.091 (0.206)
Pre Score	7.182*** (0.501)	6.364*** (0.621)	7.409*** (0.517)	3.864*** (0.733)	2.136*** (0.145)
Observations	44	44	44	44	44

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Effects of Envision treatment on employability

	(1) From a 'No' to a 'Maybe'	(2) From a 'Maybe to a Yes'	(3) From a 'No' to a 'Yes'	(4) From a 'No' to a 'Yes' or 'Maybe'
Treatment group effect	-0.0207 (0.030)	0.0510** (0.025)	0.0283 (0.025)	0.0386+ (0.0214)
Control group score	0.5850*** (0.030)	0.5716*** (0.015)	0.6529*** (0.016)	0.4489*** (0.0126)
Observations	1,192	1,680	1,494	

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

External validity

Ensuring that our results are transferable to the broader population of young people depends on the external validity of our sample. While we have covariates for age and gender on the young people in our sample, there are many measures we do not have. Additionally, it is difficult to find individual-level data among such a broad population as all young people in the UK. As such we have sought to test the external validity of our sample by comparing the mean scores in the Indices of Multiple Deprivation (IMD) for the participating schools. This meant locating the LSOAs for each of the participating schools, finding the corresponding IMD score, and then averaging between all those in our sample. Consequently this is not individual-level data, but based on the socioeconomic contextual characteristics of the schools' geographic locations.

The IMD measures seven characteristics, which taken in their aggregate feed into a single score for the level of deprivation within a geographic area. These seven characteristics are: crime, education, employment, environment, health, housing, and income.[14] A high score indicates a higher level of deprivation.

The first graph, Figure 18, shows the average difference between the aggregate IMD scores for the Lower Super Output Areas (LSOAs) of the participating schools and those of the rest of England. As can be seen, the IMD for the YSA-sample is around 12% higher than for the rest of the country, and this is a statistically significant result. The second graph, Figure 19, shows the kernel densities of aggregate IMD scores for the participating schools' LSOAs and those of the rest of the country. The peak for the non-YSA sample is negatively skewed (indicating a lower IMD). This plot confirms the above difference in means by showing that the majority of the YSA sample has an IMD score above that of the non-YSA sample (this can be seen from the blue line sitting above the orange line for more than half of the width of the plot).

[14] The IMD is produced by the Department for Communities and Local Government. The most recent round is based on 2010 data, and the seven characteristics are weighted according to DCLG's in-house specifications. For full details of these weights, and the data, see here: https://www.gov.uk/government/publications/english-indices-of-deprivation-2010

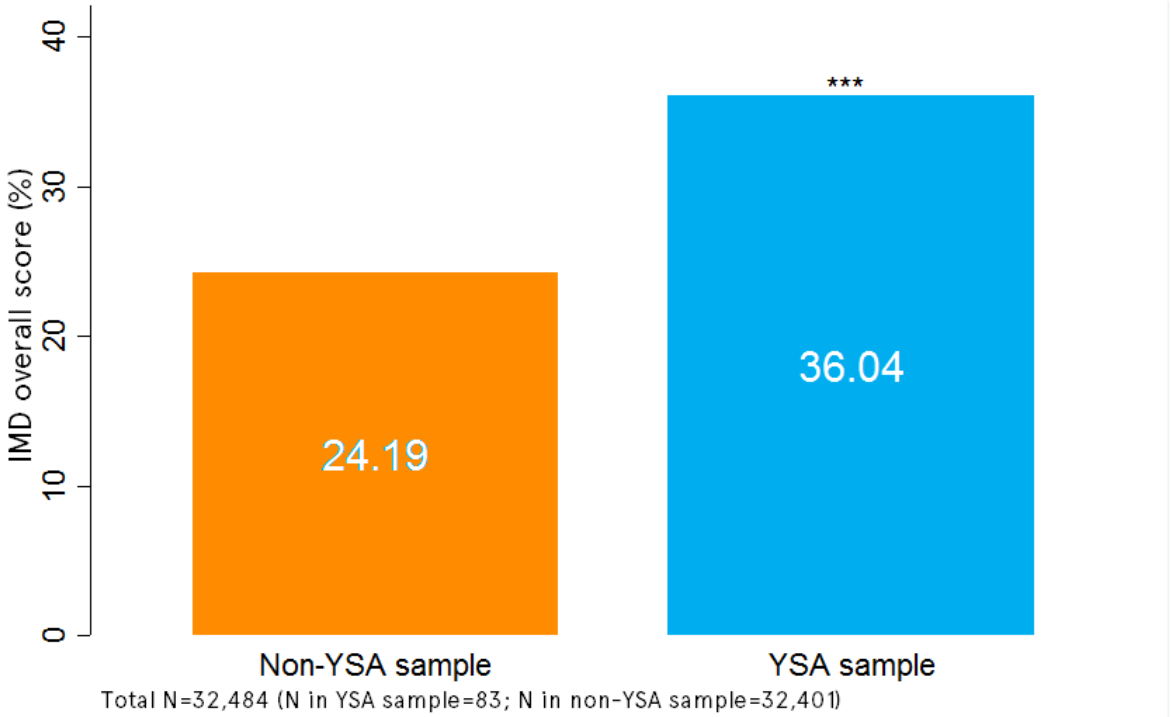


Figure 18: Mean IMD Aggregate Scores Normalised to Percentage (All English LSOAs, 2010 data)

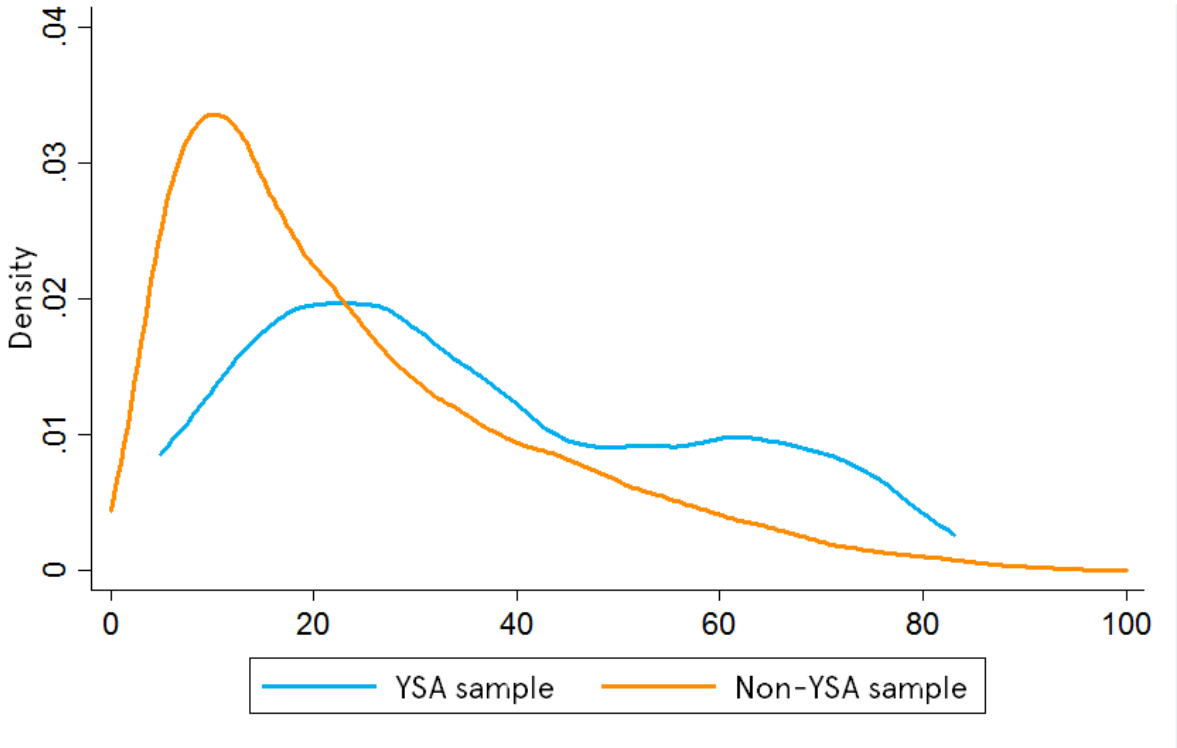


Figure 19: Kernel Densities of Aggregate IMD Scores (All English LSOAs, 2010 data)

Figures 20 to 26 show difference in means figures for each of the seven characteristics that comprise the IMD. The trend is consistent, with our sample showing a higher score in each case.

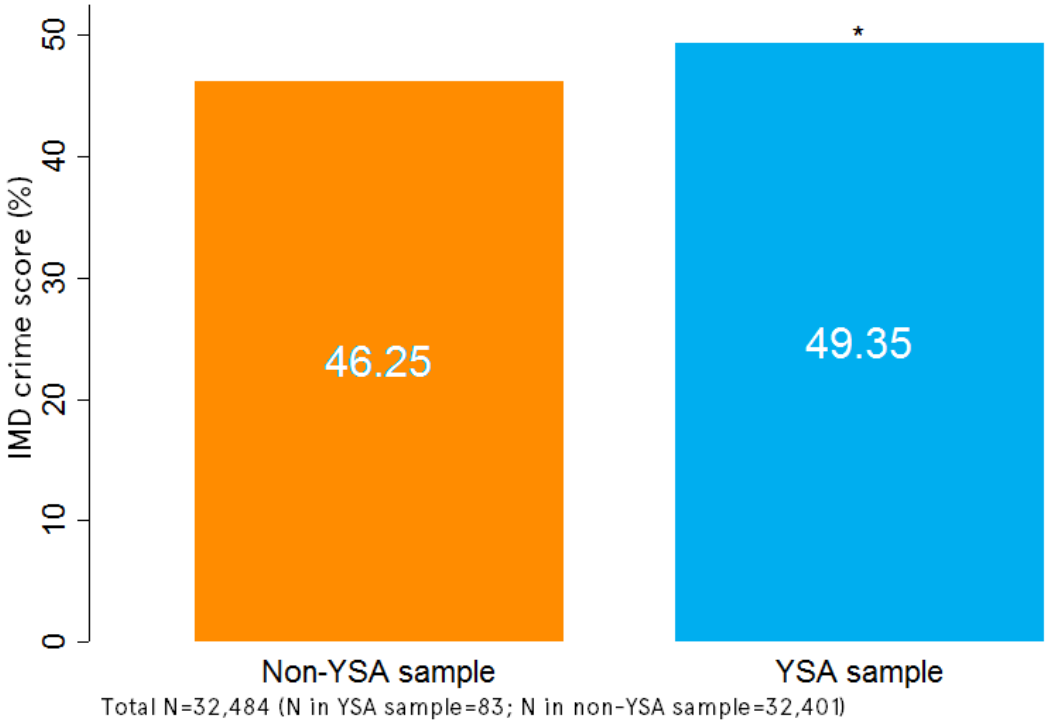


Figure 20: Mean IMD Crime Scores Normalised to Percentage (All English LSOAs, 2010 data)

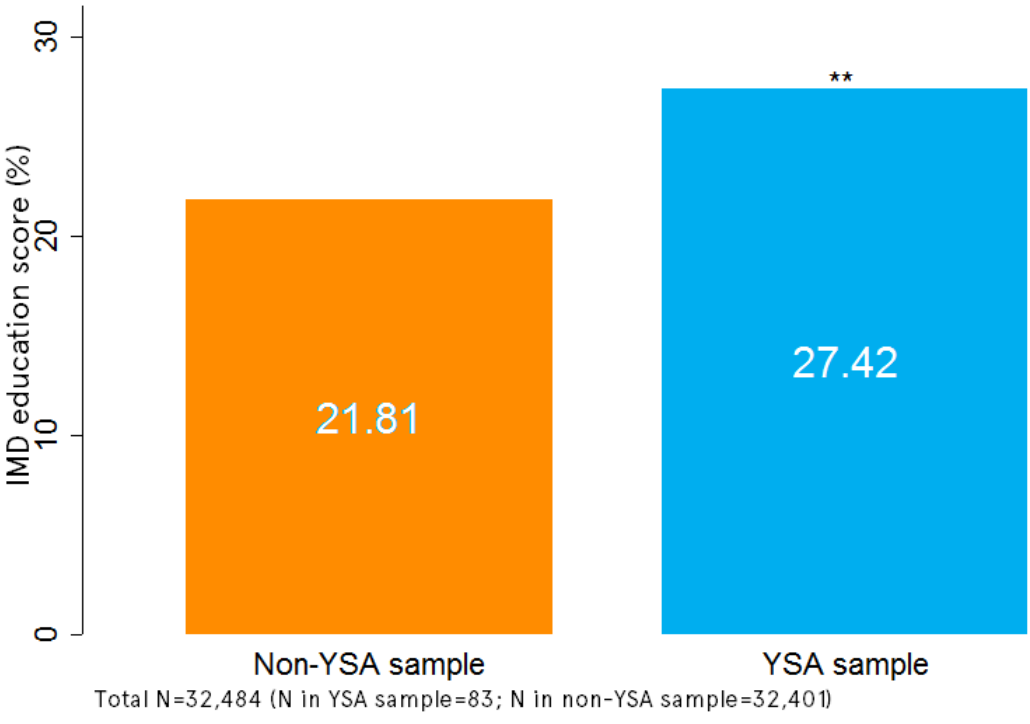


Figure 21: Mean IMD Education Scores Normalised to Percentage (All English LSOAs, 2010 data)

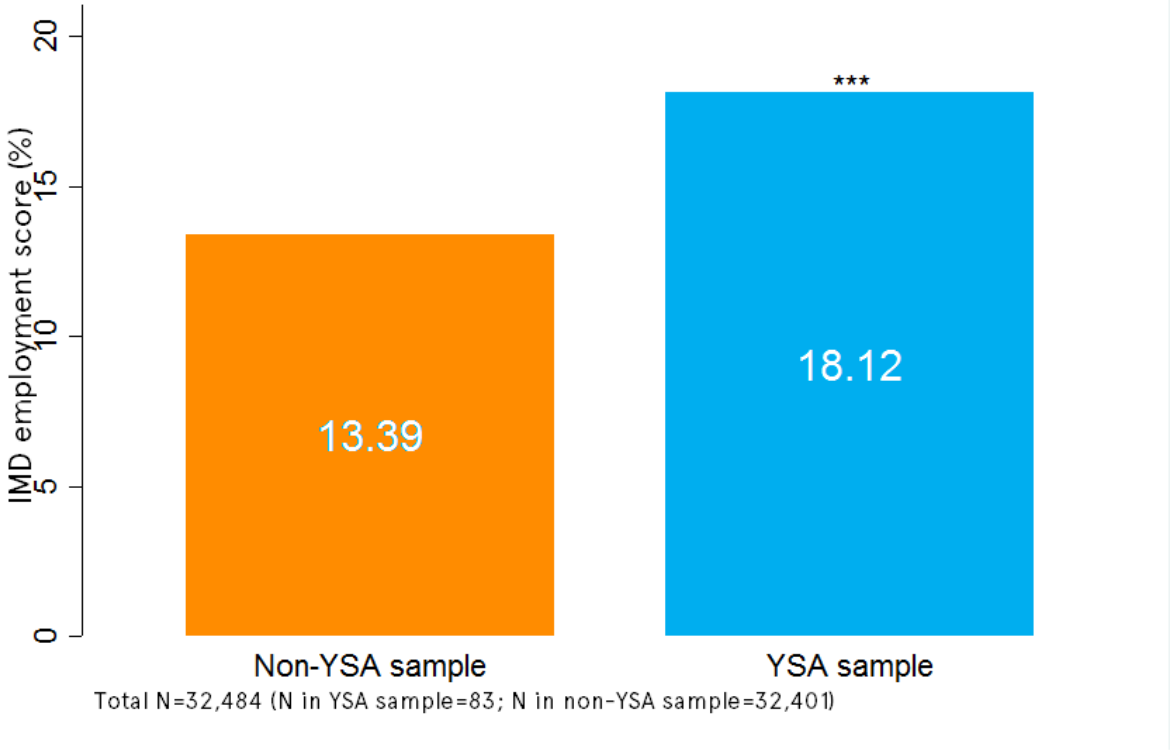


Figure 22: Mean IMD Employment Scores Normalised to Percentage (All English LSOAs, 2010 data)

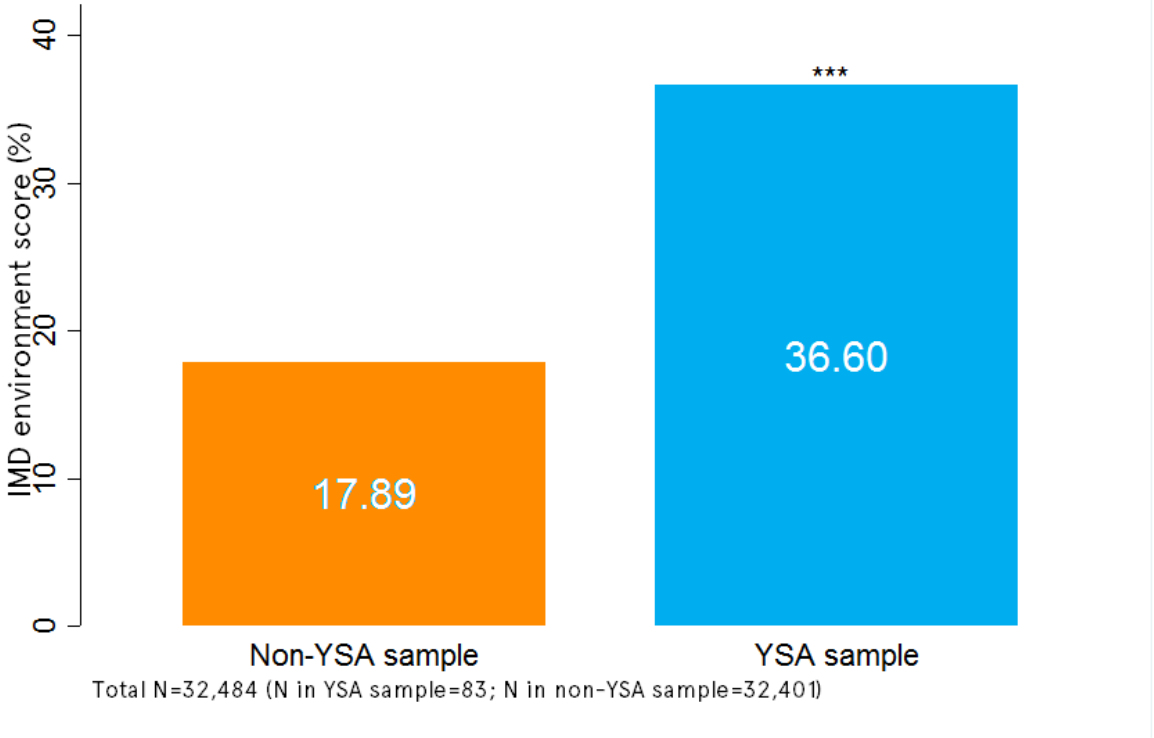


Figure 23: Mean IMD Environment Scores Normalised to Percentage (All English LSOAs, 2010 data)

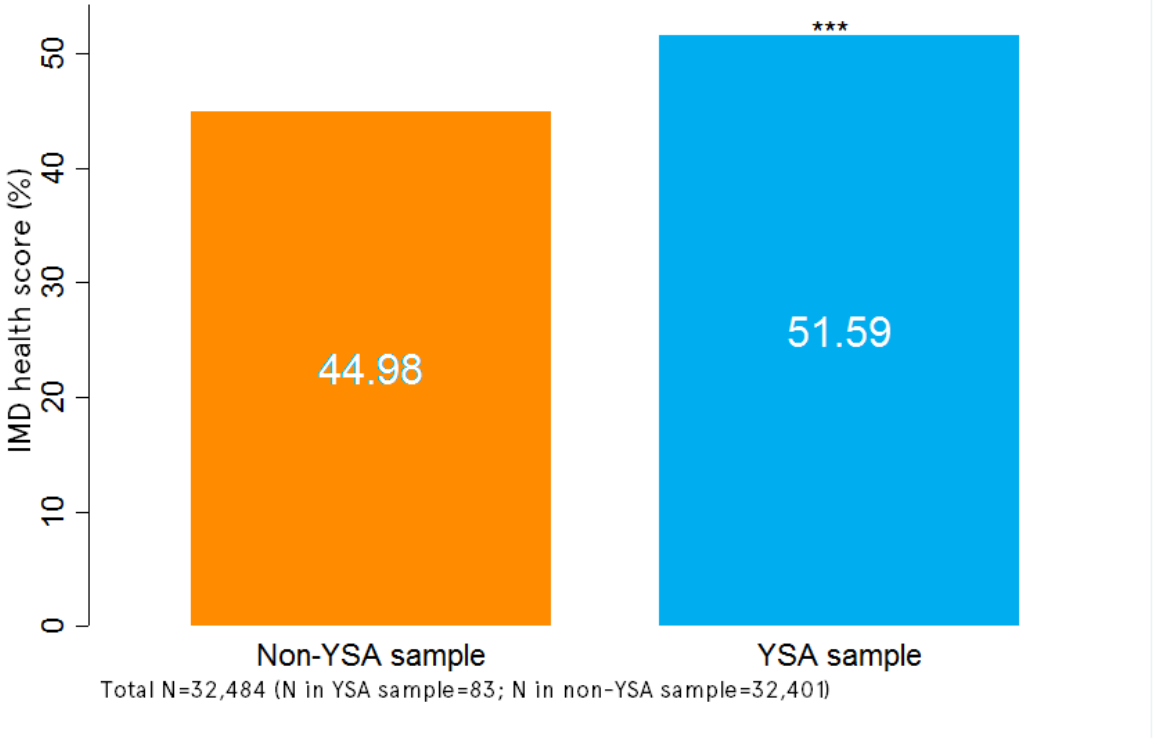


Figure 24: Mean IMD Health Scores Normalised to Percentage (All English LSOAs, 2010 data)

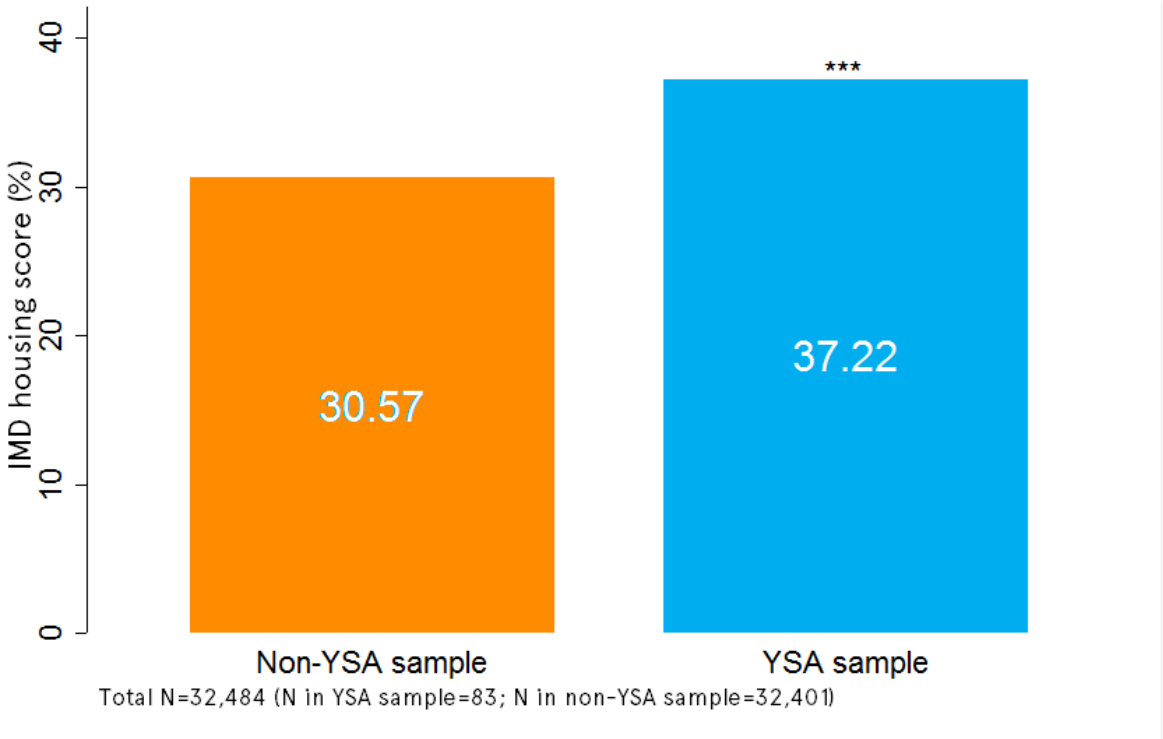


Figure 25: Mean IMD Housing Scores Normalised to Percentage (All English LSOAs, 2010 data)

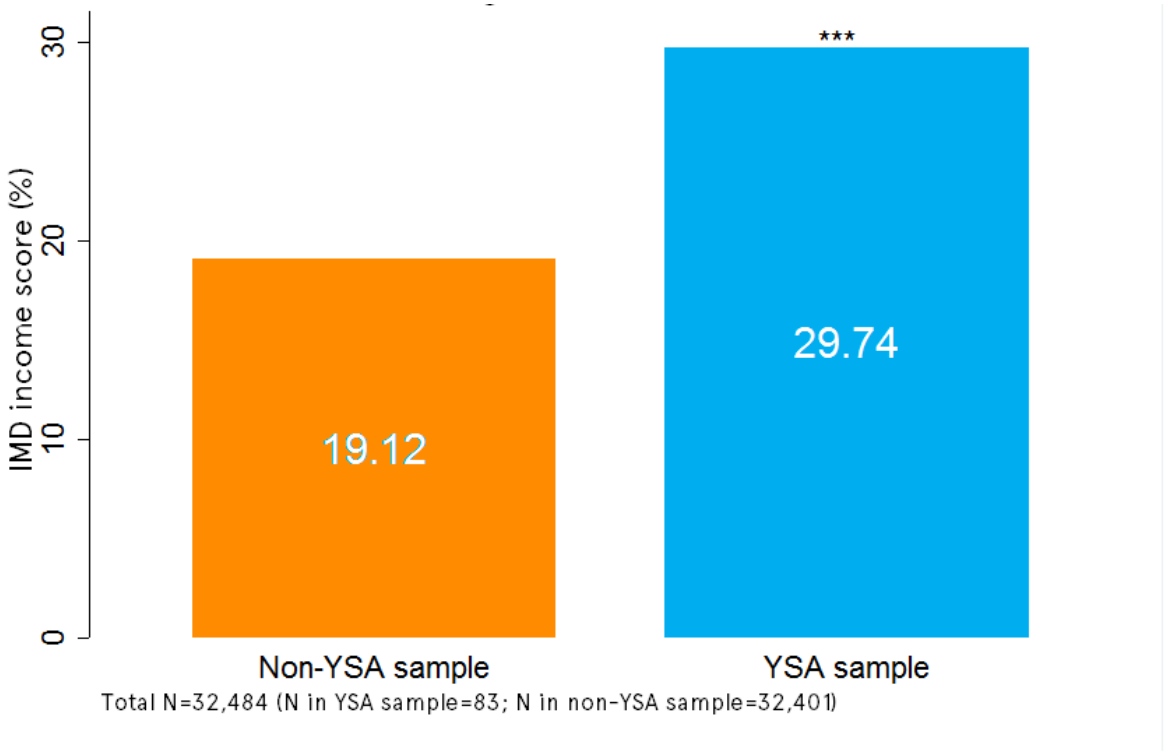


Figure 26: Mean IMD Income Scores Normalised to Percentage (All English LSOAs, 2010 data)

These figures should not necessarily be interpreted as signalling weak external validity. While our sample scores higher on every characteristic that comprises the IMD, these are not proxies for the pro-sociality measure we have been testing throughout the project. Furthermore, the IMD data is geographically based on LSOAs – which are areas covering around 50 postcodes – and is not at the individual-level such as ours. On the one hand, things such as educational level and crime are certainly important considerations when estimating levels of social capital in an area. On the other hand, as the kernel density plot above shows, there are considerable regions of overlap between our sample and the general population, for whom we can be more confident in interpreting our results.

Survey Questions

We asked the following questions to those who participated in the randomised controlled trials:

Level 0 Question	Level 1 Question detail	Response format/ options
What is your full name?	-	Text
What school year are you in?	-	Number
What is your date of birth?	-	DD/MM/YYYY
What is your gender?	-	Text
Below are some examples of volunteering activity: • Giving unpaid help at a local club, group, organisation or place of worship; • Raising money for charity; • Organising a petition or event to support a local or national issue; • Taking part in an activity to help other people or improve the local community. Over the last six months, how often have you taken part in volunteering outside of school?	At least once a week/ At least once a month/ Less often than once a month/ Never	Pick-list (one item)
How true do you think the following statements are?	**Q2_01,** I feel bad when somebody gets their feelings hurt	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q2_02, I try to understand what other people go through	Scale 0 – 10 (0 = Not true at all, 10 = Always true)

	Q2_03, I can work out my problems	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q2_04, I can do most things if I try	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q2_05, I know where to go for help with a problem	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q2_06, I am confident about having a go at things that are new to me	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q3_01, I can work with someone who has different opinions to me	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q3_02, I enjoy working together with other students my age	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q3_03, I am confident about explaining my ideas clearly	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q3_04, I am able to compromise and resolve differences of opinion	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q4_01, I often figure out different ways of doing things	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q4_02, If something goes wrong I am able to bounce back and carry on	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q4_03, Once I have started a task, I like to finish it	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q4_04, I can continue to work on things despite distractions	Scale 0 – 10 (0 = Not true at all, 10 = Always true)

	Q4_05, I am a hard worker	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q4_06, I am good at resisting temptation	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q4_07, I feel responsible for my actions	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q4_08, I feel comfortable being a group leader	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q6_01, If someone is not a success in life it's usually their own fault	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q6_02, I feel able to have an impact on the world around me	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q6_03, I feel motivated to take action on issues in my community	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q6_04, I have goals and plans for the future	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q6_05, A range of different career options are open to me	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q6_06, I'm not interested in doing any more learning	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
	Q6_07, Studying to gain qualifications is important to me	Scale 0 – 10 (0 = Not true at all, 10 = Always true)
Please answer the following:	**Q5_01,** Overall, how satisfied are you with your life nowadays?	Scale 0 – 10 (0 = Not at all, 10 = Completely)
	Q5_02, Overall, how happy did you feel yesterday?	Scale 0 – 10 (0 = Not at all, 10 = Completely)

	Q5_03, Overall, to what extent do you feel that things in your life are worthwhile?	Scale 0 – 10 (0 = Not at all, 10 = Completely)
	Q5_04, Overall, how anxious did you feel yesterday?	Scale 0 – 10 (0 = Not at all, 10 = Completely)
Generally speaking, which of the following options do you agree with most?	**Q7,** No one can be trusted/ Some people can be trusted/ Most people can be trusted/ Everyone can be trusted	Pick-list (one item only)
Are you interested in finding out more about the opportunities below in future?	**Q9,** I would like to learn more about how I can spend my own time helping people in my community in the future.	Yes/ No
	Q10, I would like to learn more about how I can spend my own time helping people in other countries in the future.	Yes/ No

How the above questions map onto the psychological constructs and/or overall wellbeing is detailed in the below table:

Aggregates	Questions for aggregates
Empathy	**Q2_01, How true (0 - 10),** I feel bad when somebody gets their feelings hurt
	Q2_02, How true (0 - 10), I try to understand what other people go through
Problem solving	**Q2_03, How true (0 - 10),** I can work out my problems
	Q2_04, How true (0 - 10), I can do most things if I try
	Q2_05, How true (0 - 10), I know where to go for help with a problem
	Q2_06, How true (0 - 10), I am confident about having a go at things that are new to me
	Q4_01, How true (0 - 10), I often figure out different ways of doing things
	Q6_02, How true (0 - 10), I feel able to have an impact on the world around me
Cooperation	**Q3_01, How true (0 - 10),** I can work with someone who has different opinions to me
	Q3_02, How true (0 - 10), I enjoy working together with other students my age
	Q3_03, How true (0 - 10), I am confident about explaining my ideas clearly
	Q3_04, How true (0 - 10), I am able to compromise and resolve differences of opinion
Grit	**Q4_02, How true (0 - 10),** If something goes wrong I am able to bounce back and carry on
	Q4_03, How true (0 - 10), Once I have started a task, I like to finish it

	Q4_04, How true (0 – 10), I can continue to work on things despite distractions
	Q4_05, How true (0 – 10), I am a hard worker
	Q4_06, How true (0 – 10), I am good at resisting temptation
	Q4_07, How true (0 – 10), I feel responsible for my actions
	Q4_08, How true (0 – 10), I feel comfortable being a group leader
	Q6_01, Agree/Disagree (1 – 10), If someone is not a success in life it's usually their own fault
Aspiration	**Q6_04, Agree/Disagree (1 – 10),** I have goals and plans for the future
	Q6_05, Agree/Disagree (1 – 10), A range of different career options are open to me
Community	**Q1,** The following list contains some examples of volunteering activities (please choose one)
	Q6_03, Agree/Disagree (1 – 5), I feel motivated to take action on issues in my community
	Q7, Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?
	Q8, We would like to make a small donation to a charity on your behalf. Below is a list of charities; please choose one.
	Q9, We are currently working with a national charity that arranges opportunities for young people to help their communities through volunteering. If you would like to be contacted with opportunities to volunteer, please tick the box below
	Q10, We are currently working with an international

	charity that arranges opportunities for young people to help people in the developing world through volunteering. If you would like to be contacted with opportunities to volunteer, please tick the box below
Education attainment	**Q6_06, How true (0 – 10),** I'm not interested in doing any more learning
	Q6_07, How true (0 – 10), Studying to gain qualifications is important to me
	National Pupil Database questions
Wellbeing	**Q5_01, Agree/Disagree (1 – 10),** Overall, how satisfied are you with your life nowadays?
	Q5_02, Agree/Disagree (1 – 10), Overall, how happy did you feel yesterday?
	Q5_03, Agree/Disagree (1 – 10), Overall, to what extent do you feel that things in your life are worthwhile?
	Q5_04, Agree/Disagree (1 – 10), Overall, how anxious did you feel yesterday?

Each question within each section is weighted equally. As an example, if a respondent scores 8 on Q2_01 and scores 10 on Q2_02, and because there are only 2 questions in this category, he/she would score (8+10)/2 = 9 for empathy. Another consideration we make is the reverse coding of some of the responses. Q6_01, for example, "*if someone is not a success in life it's usually their own fault (Agree/Disagree)*", is a negatively framed question where a response of disagreement indicates a positive outcome. In this case the scale is inverted so the highest score becomes the lowest.