TASO Transforming Access and Student Outcomes in Higher Education

TASO Summer Schools Evaluation - Interim analysis report of exploratory outcomes

November 2023

Authors: Patrick Taylor, Laure Bokobza, Pujen Shrestha - The Behavioural Insights Team

The full protocol for this study can be found on the TASO website here.

The study was pre-registered on OSF Registries: https://osf.io/tgqpa.

1. Summary

Background: This project is a collaboration between the Centre for Transforming Access and Student Outcomes in Higher Education (TASO), five Higher Education Providers (HEPs) and the Behavioural Insights Team (BIT). In summer 2022, a series of summer schools were delivered with the aim of widening participation in higher education (HE) among participants. Three types of evaluation are being conducted with these summer schools: an impact evaluation, a cost evaluation, and an implementation and process evaluation (IPE). This report presents the interim findings from the impact evaluation.

Aims: The aim of the project is to investigate the efficacy of summer schools as a widening participation activity. The aim of the widening participation agenda is to increase progression to HE among students from disadvantaged or under-represented groups.

Intervention: This study evaluated a collection of interventions. Five HEPs delivered their own summer schools, either for students in pre-16 or post-16 education. Mode of delivery differed, with some summer schools taking place in-person, and some using a combination of online and in-person elements.

Design: This study is a two-arm, parallel group randomised controlled trial (RCT).

Outcome measures: The outcomes analysed in this interim report are survey measures of participants' self-reported applications to HE, and self-reported attitudes to HE, covering their likelihood of going on to further academic study (for pre-16 students), their self-efficacy relating to HE, the compatibility of HE with their social identity, and their perception of practical barriers to HE.

Analyses: A combination of logistic and Ordinary Least Squares (OLS) regressions are used, as appropriate, to estimate effects on the primary, secondary and exploratory outcomes.

Results: There is a positive effect on students' sense of the compatibility of HE with their identity, which is significant at the 5% or 10% level depending on the model specification. Results are not highly consistent across model specifications. None of the other effects are statistically significant at the 5% or 10% level. A high and differential rate of attrition has led to a small sample and possible bias in some of the estimated effects.

Conclusions: There is early evidence of promise that these summer schools had a small positive effect on the hypothesised mediating mechanism of compatibility of HE with social identity. The analysis also suggests that there was no effect on self-reported applications to HE, self-efficacy relating to HE, students' self-reported likelihood of attending HE or post-16 academic study (depending on their age), or perception of practical barriers to HE. This is probably because most applicants to HE summer schools already intend to apply to HE. The more robust test of the intervention will come in 2025 when we have administrative data on students' entry to HE.

2. Introduction

2.1. Background

This project is a collaboration between TASO, five HEPs, and BIT. In summer 2022, a series of summer schools were delivered with the aim of widening participation in HE among participants. Three types of evaluation are being conducted with these summer schools: an impact evaluation, a cost evaluation, and an IPE. This report presents the interim findings from the impact evaluation of these summer schools. This is an extension of another research project investigating the impact of university summer schools. The previous project evaluated a collection of summer schools that were delivered online in summer 2021 due to COVID-19 restrictions. The current project evaluates summer schools that took place in 2022 in-person, though some had online elements. Roles and responsibilities for this evaluation are as follows.

BIT was responsible for:

- design, analysis and reporting for the impact evaluation;
- randomly assigning participants to the treatment or control group for the impact evaluation;
- design, data collection, analysis and reporting for the cost evaluation; and
- collecting the university entry data from the Higher Education Statistics Agency (HESA) and covariate data from the National Pupil Database (NPD).

TASO was responsible for:

- collecting all data for the impact evaluation (except for NPD and HESA data), from HEPs, from participants directly through online surveys, and through the Higher Education Access Tracker (HEAT), and;
- designing and implementing the IPE.

The five HEPs were responsible for:

- delivering the summer schools;
- collecting registration data from summer school applicants; and
- participating in the impact evaluation, IPE and cost evaluation.

HEPs were provided with funding from TASO for the project and in some cases used this to recruit a research assistant (RA) to support them with their evaluation responsibilities. In other cases, existing staff in the evaluation or widening participation teams supported the project. The table below summarises the key project personnel for each organisation.

Organisation	Name	Role and responsibilities
BIT	Dr Patrick Taylor	Principal Investigator and Evaluation QA
	Dr Laure Bokobza	Evaluation Manager

	Pujen Shrestha	Data Analyst
	Ruth Persian	Evaluation QA
TASO	Dr Helen Lawson	Research Programme Manager. IPE Lead and responsible for the day-to-day management of the study.
	Sarah Chappell	Research Manager. RCT Lead and responsible for supporting the team in the day-to-day management of the study.
	Dr Eliza Kozman	Deputy Chief Executive. Responsible for overseeing the implementation of the study.
University of Kent	Amy Burt	Project Lead at the University of Kent. Responsible for implementing randomisation and data collection there.
	Kritty Treebhoohun	RA supporting data collection and analysis.
Nottingham Trent University (NTU)	Laura Hope	Project Lead at NTU. Responsible for implementing randomisation and data collection there.
	Peter Cassidy	Co-project Lead.
University of Gloucestershire	Liz Gray	Project Lead at the University of Gloucestershire. Responsible for implementing randomisation and data collection there.
	Hannah Kent	RA supporting data collection and analysis.
University of Leeds	Liz Hurley	Project Lead at the University of Leeds. Responsible for implementing randomisation and data collection there.
	Blagovesta Tacheva	RA supporting data collection and analysis.
University of Leicester	Dr Charlotte Barratt	Project Lead at the University of Leicester. Responsible for implementing randomisation and data collection there.
	Meghann Jones	RA supporting data collection and analysis.

The project is funded by TASO, and TASO is funded by the Office for Students (OfS), the independent regulator of HE in England.

2.2. Aim

The aim of the project is to investigate the efficacy of summer schools as a widening participation activity. The aim of the widening participation agenda is to increase progression to HE among students from disadvantaged or underrepresented groups. There is currently limited evidence on this topic.

A recent review commissioned by TASO found evidence of positive correlations between summer school participation and confidence and aspirations, but mixed effects on applications and entry to HE (Robinson & Salvestrini, 2020, pp.32-34). The review also noted the limited quality of the current evidence, with most existing studies using no comparison group.

The two studies identified in this review that did use comparison groups did not do so robustly; for example, comparing participants of summer schools with failed applicants, or with young people who had not applied at all (Hoare & Mann, 2011, p.1).

The one UK-based RCT of university summer schools identified found no effect on participants' likelihood of application to HE, though the sample size for this study was small and attrition was high (Bowes et al. 2019, p.57).

An evaluation of eight summer 'bridge programs' in the US, that used an RCT design, found positive effects on the pass rates of first year college maths and writing courses (Barnett et al., 2012). However, it found no effect on course participation (the number of credits earned or attempted) and no effect on persistence at college. The sample for this study was also different in important ways to the population of interest in the current evaluation. In the US study, the sample was made up of young people who had recently graduated from high school, 100% of whom had the intention of attending college at the end of the summer. The present evaluation is focusing on young people who are not as close to participation in HE; a pre-16 cohort who have not yet taken their GCSEs (let alone applied to university), and a cohort who are in their first year of post-16 education.

In summary, there is currently no strong evidence on the causal effects of this type of summer school on widening participation. This present study aims to begin to fill this gap, by answering the following questions. Among disadvantaged or under-represented groups, what is the effect of summer schools on:

1. entry to HE (the primary outcome)?;
2. entry to the HEP that delivers the summer school (the secondary outcome)?[1]

Exploratory analysis has also been prespecified to estimate the effect of summer schools on three proximal outcomes and three potential mediating mechanisms. **This interim report presents the findings on these exploratory outcomes** (described in section 3).

To answer these questions, outcomes are compared between the participants in the trial summer schools (the treatment group), and eligible applicants who were not selected to participate (the control group). The eligibility criteria applied by HEPs has ensured that the trial sample is composed mostly of disadvantaged or under-represented groups (see section 3 for more detail on this).

[1] To support the IPE, effects were also estimated for a range of potential mediating mechanisms, helping to answer the question of how any effects on the primary and secondary outcomes are created.

2.3. Intervention

2.3.1. Introduction

This study evaluated a collection of interventions. Participating HEPs delivered their own summer schools, either for students in pre-16 or post-16 education. Each summer school had its own specific characteristics, but all had the same broad aims and involved similar activities. Below, we present TASO's brief descriptions of the pre-16 and post-16 programmes. 'Appendix I: Intervention descriptions by HEP' contains a description of each summer school, broken down by provider.

2.3.2. Pre-16 summer schools

These summer schools were focused on Year 9 or Year 10 students from underrepresented/disadvantaged backgrounds to help them decide whether HE is the right option for them. They also allowed students to experience different university subjects to discover what subject options exist outside their current school curriculum. The experience generally lasted from 2-4 days, with students staying in campus accommodation. Students experienced a range of sessions including subject tasters, student life, student finance, study skills, campus tours, and evening social activities. They also had the opportunity to work with, and ask questions of, current students at the university, either in small groups or via one-to-one mentoring.

2.3.3. Post-16 summer schools

These summer schools aimed to support students in Year 12 (or the first year of post-16 education) from underrepresented/disadvantaged backgrounds in their future decisions, including whether university is the right path for them and what subject they could study. Students stayed in campus accommodation for 2-4 days and had tours of the university campus. Students also experienced subject tasters, unless the summer school they have applied to was focused on one subject stream in particular, and were usually required to complete a project or assignment in the subject area of their choice. Other sessions aimed to give students more information and guidance on university including student finance, how to apply to university, how to write a good personal statement and choosing a university and course.

Summer School	Target group	Mode of delivery
University A	Pre-16	In person
University D Pre-16	Pre-16	In person
University D Post-16	Post-16	In person
University E	Pre-16	In person

University B (Biosciences)	Post-16	In-person and online elements
University B (Dentistry)	Post-16	In person
University B (Medicine)	Post-16	In person
University B (Psychology)	Post-16	In-person and online elements
University B (Social Sciences)	Post-16	In person
University C (Arts)	Post-16	In person
University C (Business)	Post-16	In person
University C (Law)	Post-16	In person
University C (Medicine)	Post-16	In person
University C (STEM)	Post-16	In person

Table 1: Summary of summer school delivery

3. Methods

3.1. Design

This study is a two-arm, parallel group randomised controlled trial (RCT), testing for superiority of the treatment condition over the control condition. Eligible applicants to the summer schools were randomly assigned to either the treatment or control group. Each summer school programme had a different number of places available, a different number of eligible applicants, and a different set of quotas that HEPs wished to fulfil in their participant pool, so the ratio of assignment differed by programme. See 'Randomisation' below for details of the assignment procedure.

Study activities took place from January 2022 and will run until December 2025[2] (including final reporting).

3.2. Randomisation

3.2.1. Introduction

Four practical constraints were imposed by the programme that affected the randomisation:

i. Some HEPs guaranteed places for applicants meeting certain criteria (e.g. care leavers)
ii. Some HEPs had quotas that they want to fill in the treatment group (for example, a 50/50 male-female split), and these quotas varied by HEP (see Appendix III for full quota details);
iii. Applicants had to be randomised in batches to account for the rolling basis on which HEPs shared eligible applicants data with TASO; and
iv. It was possible that some students applied to more than one summer school.

These constraints added complexity to the randomisation, so the detailed step-by-step process that was followed is provided below. Randomisation was conducted at the individual level and was blocked, with the block influencing the probability of assignment. The characteristics of the blocks were defined by each summer school, based on the characteristics of their applicant pools. Individuals in the same block had the same probability of assignment. As randomisation was conducted within blocks (and not across blocks), this was a stratified randomisation, in which each block was a strata. The randomisation strategy differed from a standard stratification strategy in that we did not randomly allocate half of the candidates to the treatment and half to the control group. Instead we allocated the required number of candidates to the treatment group (corresponding to the available summer school places) and the remainder to the control group. The differences in probabilities of assignment between blocks are accounted for in the analysis by including a categorical control variable in the regression model that indicates the individual's block (block fixed effects).

[2] This is an estimate based on BIT getting access to the linked HESA-NPD data in June/July 2025 as planned.

3.2.2. Randomisation procedure[3]

TASO provided BIT with a series of spreadsheets containing a list of all eligible applicants for each individual summer school. The variables that were used for randomisation were as follows:

- First name
- Last name
- Name of summer school
- Participant ID
- Sex (M/F)
- Guaranteed place (Y/N)
- School provider

Randomisation was conducted in R, with each batch of the randomisation being quality assured by a separate researcher at BIT before the final randomised dataset was sent to TASO, who shared it with the relevant HEP. Applicants were allocated to treatment/control conditions on a rolling basis in five batches, as follows.

First batch

For this batch we:

1. Appended applicant lists from different summer schools.
2. Assigned guaranteed places. All applicants with a characteristic that guaranteed them a place were assigned to participate in the summer school, *but not included in the trial analysis*.
3. For each applicant that applied to more than one summer school in the batch, we randomly selected which summer school they were considered for, using a random number generator. We created a variable (ENTERRAND) taking value 1 if the applicant entered randomisation for that summer school, 0 otherwise. This strategy was used to ensure that if two applicants in the same batch applied to the same set of summer schools, they could not be selected to participate in the randomisation for the same summer school.
4. For each summer school in the batch, we assigned applicants with ENTERRAND = 1 to treatment/control. This was done as follows.
    a. We split the applicant list according to the quota variable (e.g. sex). Using the 50/50 sex quota as an example, we assigned females a computer-generated random number.

[3] This section describes the randomisation procedure for the trial. Eligible applicants who did not consent to participation in the trial were also randomly assigned to either participate in the summer school that they applied to or not. This affected the number of places available in the treatment group for the trial. BIT carried out this randomisation before randomising consenting applicants into the treatment or control group for the trial. HEPs shared the study IDs of non-consenters with TASO (who shared these with BIT) for the purposes of this randomisation procedure.

b. We sorted the random numbers in ascending order.
c. We allocated 50% of the available places to the corresponding number of female applicants at the top of the list. For example, if there are 30 places available in total (after having subtracted the guaranteed places), the first 15 female applicants on the randomly sorted list received a place at the summer school.
d. We allocated all remaining female applicants to the control group.
e. We repeated steps (a) to (d) to allocate the remaining 50% of places available to males on the list.

The subsequent batches of randomisation were conducted as follows.

1. We checked if any applicants appeared in a previous batch using a unique ID. If so, we assigned ENTERRAND=0 to the applicant for the summer schools in the current batch (so that they could not be assigned to either the treatment OR control group in this batch). This did not apply to participants with guaranteed places, who were given places in all summer schools to which they applied (and for which they met the criteria for being guaranteed a place).
2. We repeated steps 3 and 4 outlined above for the first batch.

The strategy used means that the order in which a batch entered the randomisation process affected the number of students who could enter the randomisation for each summer school. In later batches, every applicant who applied to a summer school in a previous batch was automatically excluded from entering randomisation.

Trial participants and intervention implementers were not blind to assignment. Intervention deliverers and participants were informed about the study. They knew who had been assigned to the treatment group because they were delivering or receiving the only intervention being tested.

3.3. Outcome measures

The outcomes being measured in this trial are described in Table 2. They are broken down into three categories: primary, secondary and exploratory, defined as follows.

- **Primary outcome:** The main change that the intervention is trying to make.
- **Secondary outcomes:** The other changes that the intervention is trying to make, that are also considered to be valuable ends in themselves.
- **Exploratory outcomes:** There are two types of exploratory outcome in this study:
    - *Proximal outcomes:* Short-term indicators of primary or secondary outcomes.
    - *Mediating mechanisms:* Intermediate changes that explain how the intervention causes the primary or secondary outcomes, that are not considered to be valuable ends in themselves (distinguishing them from secondary outcomes).

These definitions are used here to help clarify the intervention's theory, but also to determine some important analytic choices. The primary outcome was used as the basis for power calculations and the primary/secondary/exploratory distinction is used to make choices about adjustments for multiple comparisons. The headline findings of the impact evaluation are the estimated effects on the primary and secondary outcomes. **This interim report covers the exploratory outcomes (highlighted in green in Table 2).** Data on these outcomes was gathered from two surveys: one administered at the end of summer school delivery ('survey 1') and one in January 2023, after the equal considerations window for HE applications via the Universities and Colleges Admissions Services (UCAS) had closed ('survey 2').

The sample is made up of two different age groups (those in pre-16 education and those in post-16 education). Not all outcome data is available for both cohorts. The final column of Table 2 indicates which cohort the relevant data is available for and, therefore, defines the sample for analysing each outcome.

Outcome measure	Data to be collected	Aggregation of items	Point of collection	Sample
PRIMARY: Progression to HE	Does the individual enter HE in the academic year 2022/23 according to the HESA dataset? Binary: yes/no	NA	After endpoint (May 2025)	Post-16 only
SECONDARY: Progression to host university	Does the individual go on to study at the HEP that delivers the summer school applied to according to the HESA dataset? Binary: yes/no	NA	After endpoint (May 2025)	Post-16 only
EXPLORATORY 1 (PROXIMAL): Application to HE	Survey 2: Have you applied to university? Binary: yes/no	NA	After endpoint (January 2023)	Post-16 only
EXPLORATORY 2 (PROXIMAL): Likelihood of going to HE	Survey 1: How likely are you to apply to university? Likert: 7-point "Extremely likely to extremely unlikely"	NA	Baseline After endpoint (Aug and Sept 2022)	Both
EXPLORATORY 3 (PROXIMAL): Likelihood of progressing to academic study post-16[4]	Survey 1: How likely is it that you will study at school or a sixth form after you've finished Year 11? Likert: 5-point "Extremely likely to extremely unlikely"	NA	Baseline After endpoint (Aug and Sept 2022)	Pre-16

[4] This is a short-term indicator of a secondary outcome (actual progression to academic study), but the latter will not be measured as part of this study as it falls outside of the study timeline.

EXPLORATORY 4 (MEDIATOR): Self-efficacy relating to HE	Survey 1: 1. How confident are you that you could make a successful application to university? 2. How confident are you that you could succeed at university? Likert: 5-point "Extremely confident" to "Not confident at all"	Mean average	Baseline After endpoint (Aug and Sept 2022)	Both
EXPLORATORY 5 (MEDIATOR): Compatibility of HE with social identity	Survey 1: How much do you agree with the following: "University is for people like me"? Likert scale: 5-point "strongly agree to strongly disagree"	NA	Baseline After endpoint (Aug and Sept 2022)	Both
EXPLORATORY 6 (MEDIATOR): Perception of practical barriers to HE	Survey 1: 1. How confident are you that you could afford to go to university? 2. How confident are you that you know how to apply to university? Likert: 5-point "Extremely confident" to "Not confident at all"	Mean average	Baseline After endpoint (Aug and Sept 2022)	Both

Table 2: Outcome measures

3.4. Sample selection

The study sample was made up of all applicants to the trial summer schools who met the HEPs' eligibility criteria. These criteria varied slightly by HEP, but the following list covers all criteria used across providers in the study. To have been eligible for consideration, an applicant must have had some combination of the following characteristics:[5]

- identify as coming from a Black or minority ethnic background;
- identify as Gypsy, Roma, or Traveller;
- live in an area of deprivation as defined by the most deprived quintile (Q1) of the Index of Multiple Deprivation (IMD);
- live in an area of low participation in HE as defined by the POLAR classification (Q1 or Q2);
- be care-experienced;
- be estranged (students who have lost contact with their parents and/or are studying without the support of their parents);
- be a young carer;
- have a disability;
- be the first in their family to attend HE;
- be eligible for free school meals (FSM);

[5] For some summer schools, if an applicant has one or more of the following characteristics, they were guaranteed a place on the summer school, so will not be randomised and become part of the study sample: a care leaver, care-experienced, live in a low participation area as defined by POLAR. Appendix III gives a full breakdown by summer school of the characteristics that guaranteed applicants a place.

- indicate an interest in a subject offered by the HEP;
- indicate an interest in studying close to home;
- have a low household income (£25,000 per annum or below);
- have had their studies disrupted by circumstances in their personal, social or domestic lives (for example, through trauma, medical or mental health issues); and/or
- be a refugee or asylum seeker.

University A, D, and E all had the additional criterion that all applicants must attend a school that partners with the university. This means that only students who attended a partner school were deemed eligible for the summer school. Partner schools have a longstanding relationship with the university based on their location and high percentage of students that meet widening participation criteria.

The sample was divided into two age groups: a pre-16 and post-16 group. The pre-16 group contained individuals from Years 9 and 10. The post-16 group contained individuals from Year 12 or in the first year of post-16 education.

Recruitment of study participants was carried out by the HEPs who agreed to participate in the trial. The size of the sample was determined by the number of eligible applicants to the summer schools run by these HEPs. The size of the treatment group was determined by the number of places available in each summer school.

3.5. Analytical strategy

3.5.1. Exploratory outcome 1

The following model was used to estimate the effects of the intervention on exploratory outcome 1. Analysis was conducted on an intention-to-treat basis, including all complete cases in the post-16 sample.

$$Y_i \sim bernoulli(p_i)\ ;\ logit(p_i) = \beta_0 + \beta_1 T_i + \beta_2 X_i$$

where,

- Y_i is a binary indicator of **whether the individual had applied to HE by January 2022 - self-reported** (1 if they have, 0 if not);
- p_i is the probability of Y_i;
- T_i is binary indicator of treatment assignment (1 for treated, 0 for control); and
- X_i is a vector of pre-treatment covariates (summer school applied to, sex, ethnicity, postcode-level marker of disadvantage, FSM status, whether anyone in the family

has been to university, academic year group, school ID, KS4 attainment 8 score, and an indicator of the block from which the individual was randomised).[6]

3.5.2. Exploratory outcome 2

The following model was used to estimate the effects of the intervention on exploratory outcome 2, using ordinary least squares (OLS) regression. Analysis was conducted on an intention-to-treat basis, including all complete cases across both cohorts.

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

where,

- Y_i is the **self-reported likelihood that the individual will apply to HE** (the score on a 7-point Likert scale);
- T_i is binary indicator of treatment assignment (1 for treated, 0 for control);
- X_i is a vector of pre-treatment covariates (summer school applied to, sex, ethnicity, postcode-level marker of disadvantage, FSM status, whether anyone in the family has been to university, academic year group, school ID, combined KS2 Maths and English score, and an indicator of the block from which the individual was randomised);[7] and
- ϵ_i is the heteroskedasticity robust residual error term.

3.5.3. Exploratory outcome 3

The following model was used to estimate the effects of the intervention on exploratory outcome 3, using ordinary least squares (OLS) regression. Analysis was conducted on an intention-to-treat basis, including all complete cases in the pre-16 sample.

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

where,

- Y_i is the self-reported **likelihood that the individual will go on to study at school or a sixth form after Year 11** (the score on a 5-point Likert scale);
- T_i is binary indicator of treatment assignment (1 for treated, 0 for control);

[6] We don't have the KS4 attainment 8 score yet, so this covariate is not included in the analysis in any of the models at this stage.
[7] We don't have the combined KS2 Maths and English scores yet, so this covariate is not included in the analysis in any of the models at this stage.

- X_i is a vector of pre-treatment covariates (summer school applied to, sex, ethnicity, postcode-level marker of disadvantage, FSM status, whether anyone in the family has been to university, academic year group, school ID, combined KS2 Maths and English score, and an indicator of the block from which the individual was randomised); and
- ϵ_i is the heteroskedasticity robust residual error term.

3.5.4. Exploratory outcome 4

The following model was used to estimate the effects of the intervention on exploratory outcome 4, using OLS regression. Analysis was conducted on an intention-to-treat basis, including all complete cases across both cohorts.

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

where,

- Y_i is the individual's **self-efficacy relating to HE** (the score on a 5-point Likert scale);
- T_i is binary indicator of treatment assignment (1 for treated, 0 for control); and
- X_i is a vector of pre-treatment covariates (summer school applied to, sex, ethnicity, postcode-level marker of disadvantage, FSM status, whether anyone in the family has been to university, academic year group, school ID, combined KS2 Maths and English score, and an indicator of the block from which the individual was randomised); and
- ϵ_i is the heteroskedasticity robust residual error term.

3.5.5. Exploratory outcome 5

The following model was used to estimate the effects of the intervention on exploratory outcome 5, using OLS regression. Analysis was conducted on an intention-to-treat basis, including all complete cases across both cohorts.

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

where,

- Y_i is the level of perceived **compatibility of HE with the individual's social identity** (the score on a 5-point Likert scale);

- T_i is binary indicator of treatment assignment (1 for treated, 0 for control); and
- X_i is a vector of pre-treatment covariates (summer school applied to, sex, ethnicity, postcode-level marker of disadvantage, FSM status, whether anyone in the family has been to university, academic year group, school ID, combined KS2 Maths and English score, and an indicator of the block from which the individual was randomised); and
- ϵ_i is the heteroskedasticity robust residual error term.

3.5.6. Exploratory outcome 6

The following model was used to estimate the effects of the intervention on exploratory outcome 6, using OLS regression. Analysis was conducted on an intention-to-treat basis, including all complete cases across both cohorts.

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

where,

- Y_i is the individual's **perception of practical barriers to HE** (a mean average of scores for this 2-item scale);
- T_i is binary indicator of treatment assignment (1 for treated, 0 for control); and
- X_i is a vector of pre-treatment covariates (summer school applied to, sex, ethnicity, postcode-level marker of disadvantage, FSM status, whether anyone in the family has been to university, academic year group, school ID, combined KS2 Maths and English score, and an indicator of the block from which the individual was randomised); and
- ϵ_i is the heteroskedasticity robust residual error term.

3.5.7. Descriptive statistics on the impact of COVID-19

Outcome survey 1 included two questions that asked respondents to consider the effect of the COVID-19 pandemic on their future plans (using 5-point Likert scales). TASO has hypothesised that this may moderate the effects of the intervention. The mean and standard deviation of the scores for these two items is reported by treatment condition to aid interpretation of the results in the IPE. A formal test for heterogeneous effects was not carried out in this case because it was not possible to recover an unbiased estimate when the moderating factor is realised post-intervention (as in this case).

4. Results

4.1. Participant flow

Figure 1 presents a CONSORT flow diagram of the trial so far, with an overview of the timings and sample numbers for recruitment, intervention delivery and follow-up. Students are considered to have participated in a summer school if they passed the threshold defined by the HEP for the compliance analysis (see section 12.13 of the research protocol for a list of these thresholds broken down by HEP). The proportion of compliers in the treatment group was 55%. This participation information is included in the flow diagram for completeness, but has not been used in the analysis for this report, which is all done on an intention-to-treat basis. A complier average causal effect (CACE) will be estimated for the primary outcome in the final report.

Table 3 and Table 4 summarise the attrition through survey 1 and survey 2 in this trial. Attrition for these surveys was high and occurred at different rates in the treatment and control groups for both surveys. Attrition was lower in survey 1 than survey 2. Attrition rates also differed across summer schools as highlighted in Table 5. The differential attrition by summer school will not have introduced any bias into the estimated effects because randomisation was stratified by summer school. Analysis in the balance checks section below reveals that the survey 2 sample is imbalanced on all observable characteristics tested. The level of attrition also means that the samples for the outcomes analysed from both surveys are small and likely underpowered (hence the wide confidence intervals on the estimated effects reported in section 4.3).

		Treatment	Control	Total
Number of Pupils	**Randomised**	661	1,026	1,687
	Analysed for outcome survey 1	341	212	553
Pupil attrition	**Number lost from outcome survey 1**	320	814	1,134
	Percentage lost from outcome survey 1	48.4%	79.3%	67.2%

Table 3: Summary of survey 1 attrition[8]

[8] We take the survey 1 sample to be the sample of students who answered at least one of the questions in that survey.

		Treatment	Control	Total
Number of Pupils	**Randomised**	487	747	1,234
	Analysed for outcome survey 2	96	110	206
Pupil attrition	**Number lost from outcome survey 2**	391	637	1,028
	Percentage lost from outcome survey 2	80.3%	85.3%	83.3%
Notes: This sample is limited to the post-16 summer schools. Only these participants were asked to complete survey 2, because they were the only ones old enough to have applied for HE at the time of surveying.				

Table 4: Summary of survey 2 attrition

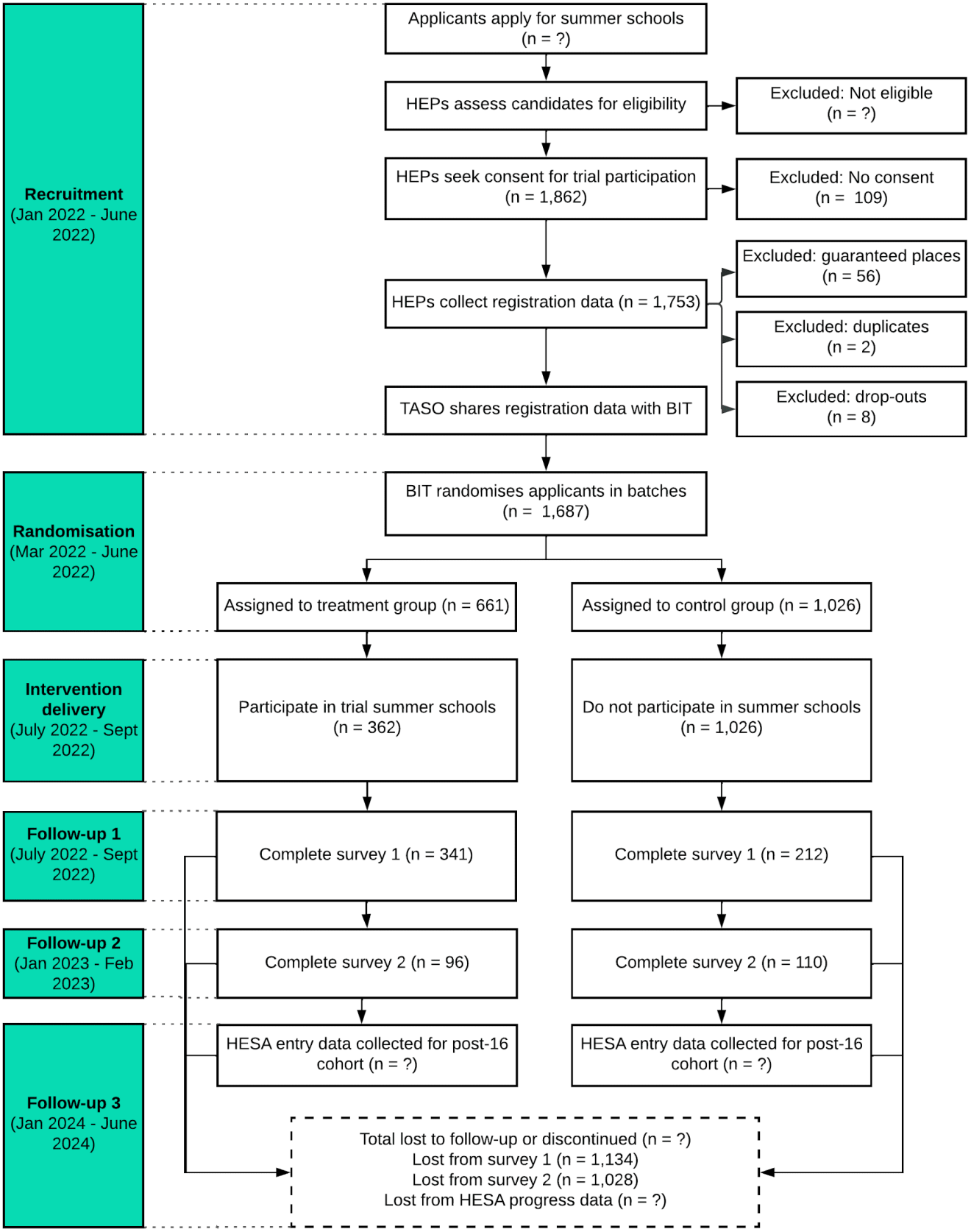

Figure 1: Study flow diagram. *Notes:* Students are considered to have participated in a summer school if they passed the threshold defined by the HEP for the compliance analysis. A '?' indicates that the data is not yet available.

	Survey 1		Survey 2		Survey Incentive
Summer School	**Response rate**	**Survey Period**	**Response rate**	**Survey Period**	
University D Pre-16	35.9%	29/7/22 – 15/9/22	—	—	£50 amazon voucher prize draw
University D Post-16	60.6%	22/7/22 – 15/9/22	39.4%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
University E	42.0%	2/8/22 – 15/9/22	—	—	£50 amazon voucher prize draw
University A	57.9%	15/7/22 – 15/9/22	—	—	£50 amazon voucher prize draw
University B (Dentistry)	37.3%	29/7/22 – 15/9/22	29.4%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
University B (Medicine)	25.2%	29/7/22 – 15/9/22	17.6%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
University B (Psychology)	27.2%	14/7/22 – 15/9/22	19.8%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
University B (Social Sciences)	14.3%	1/7/22 – 15/9/22	0.0%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
University B (Biosciences)	46.9%	7/7/22 – 15/9/22	9.0%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
University C (Arts)	42.9%	14/7/22 – 15/9/22	16.7%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
University C (Business)	32.5%	14/7/22 – 15/9/22	2.5%	25/1/23 – 24/2/23	£50 amazon voucher prize draw

University C (Law)	48.7%	14/7/22 – 15/9/22	26.3%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
University C (Medicine)	12.8%	14/7/22 – 15/9/22	15.0%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
University C (STEM)	36.4%	14/7/22 – 15/9/22	24.7%	25/1/23 – 24/2/23	£50 amazon voucher prize draw
Notes: Survey 2 was only administered to participants in post-16 summer schools.					

Table 5: Data collection and survey response rates by HEP

4.2. Description of data

Sample demographics

Table 6 shows the baseline demographic characteristics for each group in three samples: randomised participants, survey 1 responders, and survey 2 responders. The survey 1 sample is very different to the general population in three ways. Compared to the population of England at the same age, the total sample contains a higher proportion of students eligible for free school meals[9] (FSM) (29% vs 20%), a higher proportion of girls[10] (69% vs 49%), and a smaller proportion of White students[11] (47% vs 82%). These comparisons are as expected, because ethnicity was used by HEPs as a selection criterion, and a greater proportion of girls and ethnic minority students enter HE (so greater proportions would be expected to apply for HE summer schools). FSM status (along with some other indicators of socio-economic status) was a selection criterion for most summer schools so we would expect to see a higher proportion of students eligible for FSM in the sample.

The survey 2 sample is similar to the survey 1 sample. It contains a slightly higher proportion of FSM students (33%), again this is higher than that in the English population and also higher than the randomised sample. The survey 2 sample also contains a higher proportion of female students and a lower proportion of white students (74% and 33% respectively) than the population of England at the same age.

9 https://explore-education-statistics.service.gov.uk/find-statistics/free-school-meals-autumn-term/2020-21-autumn-term

10 https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2020

11 https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest

	Randomised sample (N = 1,687)		Survey 1 (N = 553)		Survey 2 (N = 206)	
	Treatment (N = 661)	Control (N = 1,026)	Treatment (N = 341)	Control (N = 212)	Treatment (N = 96)	Control (N = 110)
Eligible for FSM						
Yes	199 (30.1%)	304 (29.6%)	104 (30.5%)	60 (28.3)	38 (39.6%)	30 (27.3%)
No	419 (63.4%)	672 (65.5%)	213 (62.5%)	139 (65.6%)	52 (54.2%)	75 (68.2%)
Unknown	35 (5.3%)	39 (3.8%)	18 (5.3%)	11 (5.2%)	6 (6.3%)	5 (4.5%)
Missing	8 (1.2%)	11 (1.1%)	6 (1.8%)	2 (0.9%)	0 (0.0%)	0 (0.0%)
Sex						
Female	460 (69.6%)	761 (74.2%)	235 (68.9%)	146 (68.9%)	67 (69.8%)	86 (78.2%)
Male	186 (28.1%)	253 (24.7%)	99 (29.0%)	61 (28.8%)	27 (28.1%)	23 (20.9%)
Other	15 (2.3%)	12 (1.2%)	7 (2.1%)	5 (2.4%)	2 (2.1%)	1 (0.9%)
Ethnicity						
White	291 (44.0%)	398 (38.8%)	156 (45.7%)	102 (48.1%)	34 (35.4%)	33 (30.0%)
Asian	171 (25.9%)	296 (28.9%)	84 (24.6%)	51 (24.1%)	34 (35.4%)	37 (33.6%)
Black	138 (20.9%)	236 (23.0%)	65 (19.1%)	42 (19.8%)	21 (21.9%)	27 (24.5%)
Other	58 (8.8%)	90 (8.8%)	34 (10.0%)	16 (7.5%)	7 (7.3%)	13 (11.8%)
Missing	3 (0.5%)	6 (0.6%)	2 (0.6%)	1 (0.5%)	0 (0.0%)	0 (0.0%)
Year						
12	487 (73.7%)	747 (72.8%)	214 (62.8%)	136 (64.2%)	96 (100%)	110 (100%)
10	108 (16.3%)	188 (18.3%)	81 (23.8%)	56 (26.4%)	0 (0.0%)	0 (0.0%)
9	66 (10.0%)	91 (8.9%)	46 (13.5%)	20 (9.4%)	0 (0.0%)	0 (0.0%)
Notes: Totals do not add up to 100% due to rounding.						

Table 6: Distribution of covariates by treatment group[12]

Balance checks

Table 7 presents balance checks on FSM status, ethnicity, sex and year group on the survey 1 sample. To assess balance, the magnitude of the differences in mean scores between the two groups is calculated for each covariate.[13] Rather than reporting simple differences in means for each covariate, normalised differences are presented to aid comparison between covariates that have different units, and to facilitate comparisons across studies.

The normalised difference is defined as the difference in means between the two groups, divided by the pooled standard deviation. Normalised differences with a magnitude of 0.1 or less indicate a negligible correlation between the covariate and assignment to treatment group, which can usually be addressed through covariate adjustment in the regression (Austin 2009, p.1233), as planned here. Following this interpretation of the magnitude of differences, the survey 1 sample is well-balanced across all covariates on FSM status, ethnicity, year group, and sex.

	Treatment		Control		
	Mean	**(S.D.)**	**Mean**	**(S.D.)**	**Normalised difference**
FSM	0.31	0.46	0.29	0.45	0.05
White	0.46	0.50	0.48	0.50	-0.05
Female	0.69	0.46	0.69	0.47	0.00
Year 12	0.63	0.48	0.64	0.48	-0.03
Notes: N = 553. All variables are binary indicators, so mean averages represent proportions of the group. The 'Unknown' category in FSM is coded as missing in the dataset, so the reported means and S.D.s are of the non-missing sample.					

Table 7: Balance checks on survey 1 sample

Table 8 shows the results of the balance checks for the survey 2 sample. The survey 2 sample is imbalanced on FSM status, ethnicity and sex. This may mean that the point estimates reported in the results below are biased, though it is not possible to accurately estimate the size or direction of this bias. Females are more likely to participate in HE than

[12] We only include here a subset of the covariates for the purposes of balance checks and assessing the representativeness of our samples compared to the general population.
[13] A common alternative is to report whether differences between groups are statistically significant at a certain level of confidence (often $p < 0.05$ in the social sciences). This approach is not particularly helpful because it only tells us whether the sample is large enough to detect a difference, and leaves open the question as to whether any observed differences – and any associated bias – can be addressed through simple covariate adjustment (the approach taken in the analysis for this study) (Imbens & Rubin 2015, p.311).

males,[14] so attendance at a summer school may have a smaller effect on their future participation in HE as compared to males. If this is the case, the greater proportion of females in the control group could contribute to some downward bias.

Two points about randomisation are important to note. First, the imbalance on one observed covariate does not mean that the joint effect of all relevant covariates (many of which are unobserved) is not balanced; this quantity is unknown. Second, the purpose of randomisation is not to ensure that point estimates are unbiased by achieving perfect balance on relevant covariates. The purpose of randomisation is to ensure that the potential distribution of estimated treatment effects (reported in the results below as 95% confidence intervals) is unbiased; i.e. if we ran the experiment 100 times, the true effects would be in the 95% confidence intervals 95% of the time.

	Treatment		Control		
	Mean	**(S.D.)**	**Mean**	**(S.D.)**	**Normalised difference**
FSM	0.40	0.49	0.27	0.45	0.26
White	0.35	0.48	0.30	0.46	0.12
Female	0.70	0.46	0.78	0.42	-0.19
Year 12	1	0	1	0	—
Notes: N = 206. All variables are binary indicators, so mean averages represent proportions of the group. The 'Unknown' category in FSM is coded as missing in the dataset, so the reported means and S.D.s are for the non-missing sample.					

Table 8: Balance checks on survey 2 sample

To investigate the source of the observed imbalance, we have repeated the balance checks for the pre-attrition sample (all randomised students). Table 9 presents the results of these checks. It shows that the pre-attrition sample is well-balanced (by the definition used in this discussion) on all covariates. Comparing Table 8 and Table 9 shows that the sample was better balanced at the point of randomisation, but that students of different characteristics dropped out from survey 2 completion at different rates in the treatment and control groups, leading to the imbalance observed in Table 8. This means that the intervention could have affected both the outcomes and the type of students who completed the outcome survey. The covariate adjustment used in the analysis below will partly adjust for this. However, it is still likely that there are differences between treatment conditions in unobserved characteristics which are non-random, are not fully correlated with our observed covariates, and will therefore lead to some bias in the results. The

[14] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/843542/Publication_HEIPR1718.pdf.

compliance analysis in the final report will check whether this imbalance in the outcome data is also seen in summer school participation. We do not expect to see it in the data for the primary and secondary outcomes because these data come from administrative sources and therefore should include the full sample.

	Treatment		Control		
	Mean	**(S.D.)**	**Mean**	**(S.D.)**	**Normalised difference**
FSM	0.31	0.46	0.30	0.46	0.01
White	0.44	0.50	0.39	0.49	0.11
Female	0.70	0.46	0.74	0.44	-0.10
Year 12	0.73	0.44	0.73	0.45	0.02
Notes: N = 1,687. All variables are binary indicators, so mean averages represent proportions of the group. The 'Unknown' category in FSM is coded as missing in the dataset, so the reported means and S.D.s are of the non-missing sample.					

Table 9: Balance checks on randomised sample

Descriptive statistics for outcomes

Table 10 presents the means and standard deviations for the outcomes, broken down by treatment group. In general, it appears that both the treatment and control group performed similarly, with the treatment group responding more positively across four outcomes and the control group responding more positively on four outcomes.

Appendix II presents a more detailed breakdown of each outcome by the responses that make up the scales. This shows that across both conditions students were generally more likely to respond positively (rather than neutrally or negatively) to the survey questions. This is probably because students who apply for a university summer school are more likely to have favourable attitudes towards HE. We can also see that the self-reported rate of application to HE among the post-16 sample by January 2023 was very high in both the treatment and control group (91% and 93% respectively).

Outcome	Treatment	Control

	Mean (SD)	Mean (SD)
Likelihood of going to HE (7-point likert scale) (N = 551)	6.44 (0.94)	6.46 (0.97)
Likelihood of progressing to academic study post-16 (5-point likert scale) (N = 128)	4.47 (0.77)	4.62 (0.81)
Self-efficacy relating to HE (5-point likert scale) (N = 545)	3.98 (0.78)	4.00 (0.72)
Compatibility of HE with social identity (5-point likert scale) (N = 528)	3.89 (0.89)	3.72 (0.94)
Perception of practical barriers to HE (5-point likert scale) (N = 543)	3.20 (0.99)	3.04 (0.99)
Perception of financial barriers to HE (5-point likert scale) (N = 546)	3.19 (1.0)	3.03 (1.0)
Perception of knowledge barriers to HE (5-point likert scale) (N = 548)	3.63 (0.96)	3.55 (1.05)
Applied to HE (binary yes/no) (N = 206)	0.91 (0.29)	0.93 (0.26)
Notes: N per outcome included in brackets above. Practical barriers were split between financial and knowledge barriers in order to better understand the disaggregated effects of the two, following a request from TASO's theme working group.		

Table 10: Average outcome scores by treatment group

Descriptive statistics on the impact of COVID-19

Table 11 presents the means and standard deviations for the survey items that measure the perceived impact of COVID-19 on participants' future study and career plans, broken down by treatment group. In general, it appears that both the treatment and control group performed similarly. Across both arms and across both survey items it appears that average responses on the 5-point likert were largely neutral, suggesting that students saw COVID-19 as having no substantial effect on their future plans.

Impact of COVID-19	Treatment	Control
	Mean (SD)	Mean (SD)
The COVID-19 pandemic has made me rethink my future plans (5-point likert scale) (N = 545)	2.98 (1.23)	2.91 (1.28)
I'm worried that I may have to change my study or career plans because of the COVID-19 pandemic (5-point likert scale) (N = 545)	2.46 (1.16)	2.34 (1.11)
Notes: N per outcome included in brackets above.		

Table 11: Descriptive statistics on the impact of COVID-19

4.3. Outcome of analysis

Pre-specified analysis

Table 12 presents the estimated average effects of the summer schools on the outcomes of interest. Likelihood of going to HE was measured using a 7-point Likert scale, and all other survey 1 outcomes were measured using a 5-point scale. Whether or not a student reported applying to university by January 2023 (the survey 2 item) was measured using a binary 'yes/no' question (coded as 1 for 'yes' and 0 for 'no').

The estimated effects are based on Model 1, which was the main model pre-specified in the trial protocol. For all outcomes, it includes a series of pre-treatment covariates in the regression.[15] Results can be interpreted as follows: The mean reported likelihood of going to HE in the control group is 6.46 on a 7-point Likert scale. The estimated effect size in Model 1 is -0.05, which means that on average, and controlling for other variables in the regression, students in the treatment group scored 0.05 points lower on that scale, but this difference is statistically insignificant.[16] As another example, the mean reported compatibility with HE in the control group is 3.72 on a 5-point Likert scale. The estimated effect size in Model 1 is 0.31, which means that on average, and controlling for other variables in the regression, students in the treatment group scored 0.31 points higher on that scale, and the difference is statistically significant at the 10% level.

Effects are also presented as standardised effect sizes, to make it easier to compare between outcomes and with other studies. Figure 2 visualises the standardised effect sizes with 95% confidence intervals.

Four of the estimated effects are directionally positive and three are directionally negative. We don't report the results from the logistic regression for the 'Applied to HE' binary outcome because the pre-specified model does not converge due to some covariates perfectly predicting and separating the outcome variable.[17] We report results for this outcome from the linear specification and the logistic regression model with no covariates in Table 13. One result that is directionally positive - Compatibility of HE with social identity - is significant at the 10% level. The remaining estimates are not significant at the 5% nor 10% level. While this may partly be due to the small size of the sample, we cannot conclude with sufficient certainty that the results represent true effects as opposed to random noise.

[15] Note that school year, which was pre-specified as covariate, drops from the regression in all the models where it is included as it is fully collinear with the summer school indicator variable.
[16] Note that this estimated effect size is not equal to the difference in the unadjusted means of this outcome for the treatment and control group respectively, as reported in Table 10, as Model 1 includes covariates. The raw difference is equal to the effect size for Model 4 (no covariates), reported in Table 13.
[17] Fitted probabilities equal to 0 or 1 occur when the summer school categorical variable and/or IMD and POLAR quintile variables are included as controls in the model. The model stops converging when the school the student is from and/or the randomisation block are included as controls in the model. This risk was flagged in the trial protocol.

Outcome	Estimated effect (score on scale)	Standard error	Standardised estimated effect	P-value
	Linear regression results			
Likelihood of going to HE (7-point likert scale) (N = 541)	-0.05	0.17	-0.06	0.74
Likelihood of progressing to academic study post-16 (5-point likert scale) (N = 119)	-0.25	0.26	-0.32	0.35
Self-efficacy relating to HE (5-point likert scale) (N = 536)	-0.10	0.14	-0.13	0.49
Compatibility of HE with social identity (5-point likert scale) (N = 519)	0.31+	0.16	0.34	0.06
Perception of practical barriers to HE (5-point likert scale) (N = 534)	0.01	0.18	0.01	0.95
Perception of financial barriers to HE (N = 537)	0.03	0.18	0.03	0.88
Perception of knowledge barriers to HE (N = 539)	0.02	0.17	0.02	0.90
	Logistic regression results[1]			
Applied to HE (binary yes/no) (N = 203)	—	—	—	—

Notes: N per outcome included in brackets above.

Standardised estimated effect is calculated as Hedges' *g* for linear regressions and Cohen's *h* for logistic regression.

'Likelihood of progressing to academic study post-16' was computed for the pre-16 sample only.

'Applied to HE' was computed for the post-16 sample only.

All other effects were computed for the combined pre- and post-16 sample.

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

[1] We don't report the results from the logistic regression for the 'Applied to HE' binary outcome because the pre-specified model does not converge due to some covariates perfectly predicting and separating the outcome variable. We report results from the linear specification and the logistic regression model with no covariates in Table 13.

Table 12: Estimated effects for the outcomes of interest

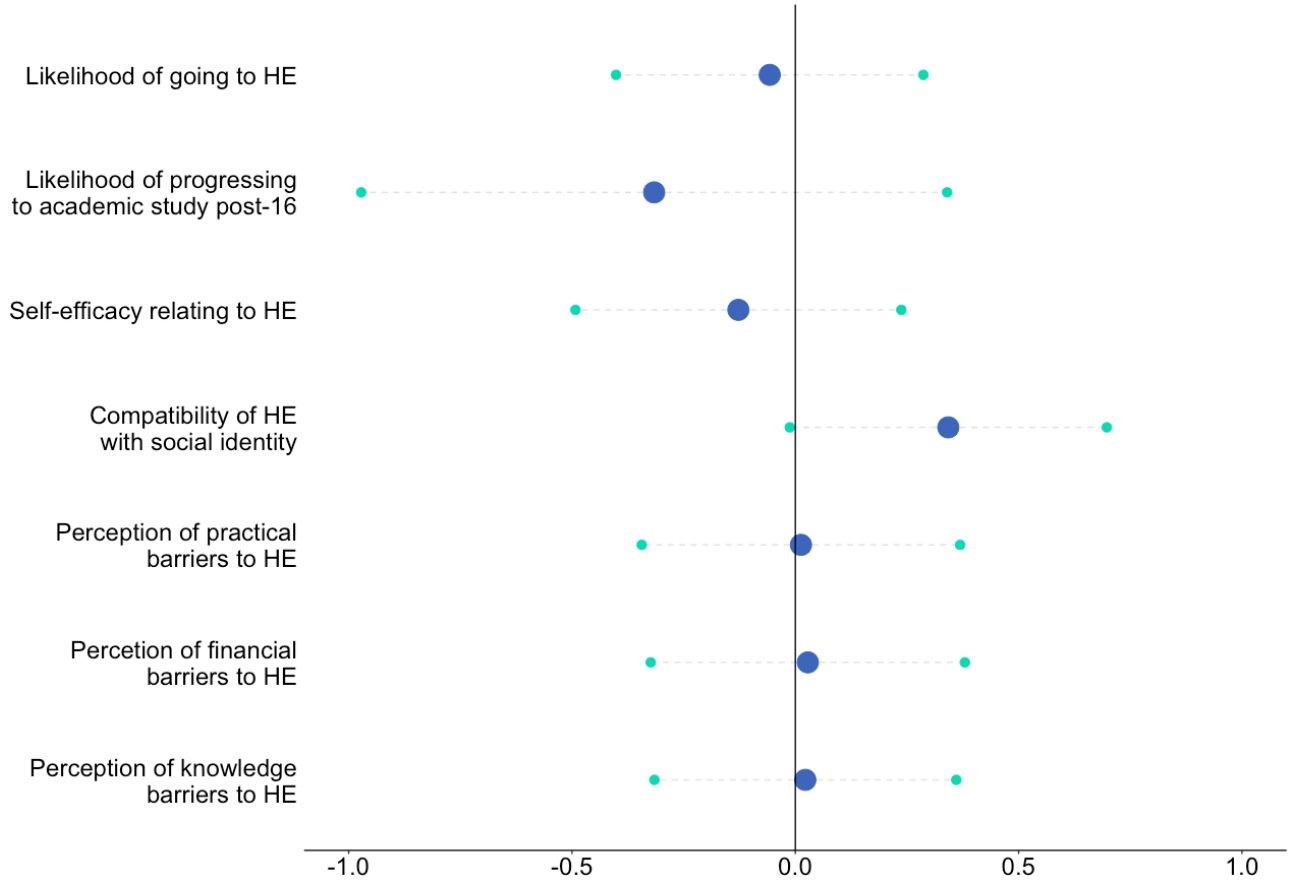

Figure 2: Estimated effect sizes in Hedges' *g with 95% confidence intervals*

Robustness checks

We have re-estimated the effects of all five outcomes for which we collected baseline data using the following three covariate combinations:

- The covariates specified in the protocol plus the baseline measure of the outcome (Model 2)
- The baseline measure of the outcome only (Model 3)
- No covariates (Model 4)

Additionally, as pre-specified we have conducted a robustness check of logistic regression estimates using linear regression (Model 5).

Lastly, we ran an additional robustness check using the missing indicator method: For the two covariates with missing values (IMD, 9 missing values, and FSM status, 19 missing values), we replace the missing value with the mean of the variable. We then include in the regression the covariate with imputed values, and an additional binary variable indicating missingness for each of the two covariates (Model 6).

Table 13 presents the estimated effects from the pre-specified model (Model 1) alongside the effects from these alternative models. It shows that, for the Likelihood of going to HE, the results from the pre-specified analysis are broadly robust to these different

specifications. The directions of all point estimates remain the same and the confidence intervals remain wide. No results are statistically significant at the 5% level.

For the Likelihood of progressing to academic study post-16 the direction of the estimates remains negative but with the model that includes the baseline measure of the outcome without any covariates, the magnitude of the effects increases and the effect becomes significant at the 10% level. It is possible therefore that the intervention has had a small negative effect on this outcome, but the sample is very small for this model and we have already noted the potential bias introduced by differential attrition.

For Compatibility of HE with social identity, we find a significant effect of the intervention at the 10% level in Models 1, 2 and 6, and at 5% in Model 4. The effect remains directionally positive across all models, but is smaller in magnitude and insignificant in Model 3, which includes the baseline outcome as control. This suggests that some level of caution should be taken when interpreting the results of the other models.

For the Perception of practical barriers to HE, the estimated effect of the intervention is small and insignificant at the 5% and 10% levels across model specifications, except in Model 4 (positive and significant at the 10% level). When looking at disaggregated barriers, we notice the same pattern for estimated effect of the intervention on financial barriers, and an insignificant effect at the 5% and 10% levels on knowledge barriers. The estimated coefficients change in direction and magnitude across the models, which suggests that they should be taken with caution.

For the Applied to HE outcome, which we were unable to estimate using the prespecified model, the logistic regression with no covariates and the linear estimate are both small and negative, but not significant at the 5% or 10% level. Full results for the linear estimation (Model 5) are presented in Appendix V.

	Estimated effects (SE)					
Outcome	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Likelihood of going to HE (7-point likert scale)	-0.05 (0.17)	-0.11 (0.17)	-0.06 (0.09)	-0.02 (0.08)	—	-0.04 (0.16)
N observations	541	469	479	551	—	551
Likelihood of progressing to academic study post-16 (5-point likert scale)	-0.25 (0.26)	-0.14 (0.33)	-0.38+ (0.20)	-0.15 (0.14)	—	-0.23 (0.22)
N observations	119	49	49	128	—	128
Self-efficacy relating to HE (5-point likert scale)	-0.10 (0.14)	-0.11 (0.14)	0.01 (0.06)	-0.03 (0.07)	—	-0.08 (0.14)

N observations	536	464	473	545	—	545
Compatibility of HE with social identity (5-point likert scale)	0.31^{+} (0.16)	0.31^{+} (0.18)	0.11 (0.08)	0.17* (0.08)	—	0.28^{+} (0.16)
N observations	519	447	456	528	—	528
Perception of practical barriers to HE (5-point likert scale)	0.01 (0.18)	-0.03 (0.16)	0.08 (0.08)	0.16^{+} (0.09)	—	0.05 (0.17)
N observations	534	462	471	543	—	543
Perception of financial barriers to HE	0.03 (0.18)	-0.03 (0.16)	0.07 (0.08)	0.16^{+} (0.09)	—	0.06 (0.17)
N observations	537	465	474	546	—	546
Perception of knowledge barriers to HE	0.02 (0.17)	-0.07 (0.19)	0.03 (0.09)	0.08 (0.09)	—	0.00 (0.17)
N observations	539	467	476	548	—	548
Applied to HE (binary yes/no)	—	—	—	-0.28 (0.51)	-0.14 (0.23)	-0.07 (0.19)
N observations	—	—	—	206	203	206

Notes:

Model 1 = model specified in protocol.

Model 2 = model specified in protocol including the outcome at baseline.

Model 3 = baseline measure of the outcome only.

Model 4 = no covariates.

Model 5 = re-estimating effects from binary outcomes using linear regression ('Applied to HE' outcome only). Note that for the 'Applied to HE' outcome, the coefficients from Model 4 and 5 are not directly comparable, since one is expressed in logit units and the other is in (percentage point / 100) units.

Model 6 = missing indicator method, imputing missing values for covariates at their mean, and adding a binary variable indicating missingness for each covariate with missing values.

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

Table 13: Estimated effects with different model specifications

5. Discussion

Interpretation

This interim analysis suggests that these summer schools may have had a small positive effect on one hypothesised mediating mechanism (students' sense of compatibility of HE with their social identity). This survey question asked students whether they feel university is for people like them and therefore, on average, the summer schools may have made students feel that they fitted in more at university.

However, none of the other effects are statistically significant at the 5% or 10% level.[18] This means that there was likely no effect on students' self-reported likelihood of attending HE or post-16 academic study, self-efficacy relating to HE, perception of practical barriers to HE, or self-reported applications to HE. If this is the case, it is probably because most applicants to HE summer schools already intend to follow these paths (as evidenced by the fact that over 90% of survey 2 respondents in both the treatment and control groups reported applying for HE by January 2023).

Generalisability

We can think about generalisability in two ways: i. the extent to which the results might be realised by other summer schools; and ii. the extent to which the results might be realised in different populations. On the first type of generalisation, it seems quite likely that the average effects achieved by the summer schools in this study would be achieved by other summer schools. This is because a range of different types of summer school were included in this study (different subjects and different approaches). These different types may be more or less effective (we are not powered to test this), but the average effects are likely to be similar across all summer schools that share similar aims and approaches.

On the second type of generalisation, we have shown that both the trial sample and the analytic samples (post-attrition) differ substantially from the general population of England in at least three important ways; the study included a much lower proportion of White young people, a higher proportion of students eligible for FSM, and a much higher proportion of girls. We would, therefore, be unlikely to observe similar average effects if the same summer schools were run with a group of students that was representative of the wider English population. Having said this, summer schools that aim to widen participation in HE would be unlikely to aim for this kind of representation. The extent to which these summer schools saw similar effects would partly depend on the extent to which their cohort of students matched the characteristics of the cohort in this study.

Trial limitations

Three issues with the study have been discussed in this report. First, only a small proportion of the total sample at least partially completed survey 1 and survey 2 (33% and 17% respectively). For each outcome, a smaller proportion still had the complete outcome and covariate data required for the analysis. This means that the study may well be underpowered to detect the effects we are trying to estimate (hence the wide confidence

[18] Except for the perception of practical and financial barriers to HE in the model with no covariates, but as explained above, these should be taken with caution given the variation in results across model specifications (Table 13).

intervals on the estimated effects). Second, some imbalance has been identified in the observed covariates for survey 2, with a greater proportion of female students in the control group as compared to the treatment group, and a smaller proportion of White students and students eligible for FSM. The balance checks suggest that this imbalance is due to some differential attrition, which is unlikely to be fully dealt with by covariate adjustment, especially where it also led to imbalance on unobservables. Females are more likely to participate in HE than males, so attendance at a summer school may have a smaller effect on their future participation in HE as compared to males. If this is the case, the greater proportion of females in the control group could contribute to some downward bias. Third, the positive result reported above, that is significant at the 10% level (students' sense of compatibility of HE with their social identity), is not robust to different model specifications.

While the estimates produced are imprecise, and there are some question marks over the validity of the results, the challenges that have led to this were expected. An outcome survey issued to students by email is unlikely to yield a high response rate, and it was also likely that certain types of student would be more likely to complete the survey (leading to differential attrition and potential bias). The intention of this interim report was to provide early evidence of the effects of the interventions, before more robust and complete outcome data becomes available. The more robust test of the intervention will come in 2025 when we have administrative data on students' entry to HE.

Bibliography

Aspire to HE: Higher education for all. (2018). *Aspire to HE Right Track Summer School—Delivery Report*. University of Wolverhampton.

Austin, P.C., 2009. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Communications in Statistics-Simulation and Computation, 38(6), pp.1228-1234.

Cohen, J., 2013. *Statistical power analysis for the behavioral sciences*. Academic press.

Church, E. (2018). *Specific Evaluation: Summer Residential*. DANCOP: Derbyshire & Nottinghamshire Collaborative Outreach Programme & EMWPREP: East Midlands Widening Participation Research and Evaluation Partnership.

Hatt, S., Baxter, A., & Tate, J. (2009). 'It was definitely a turning point!' A review of Aimhigher summer schools in the south west of England. Journal Of Further And Higher Education, 33(4), 333-346.

Hayes, C., Hudson-Jones, E., Church, E., & Burr, E. (2018). *Higher Horizons+ Unify Residential 2017 Evaluation*. East Midlands Widening Participation Research and Evaluation Partnership.

Imbens, G.W. and Rubin, D.B., 2015. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.

Lei, J., Calley, S., Brosnan, M., Ashwin, C. and Russell, A., 2020. Evaluation of a transition to university programme for students with autism spectrum disorder. Journal of Autism and Developmental Disorders, 50(7), pp.2397-2411.

Appendix I: Intervention descriptions by HEP

University A

A four-day on-campus summer school for Year 10 students with three nights spent in university accommodation. Activities are split into four sections: information on HE, subject specific, social, and student life. In the four-day period, 35 sessions were offered consisting of four HE information sessions, six subject tasters, four student life activities and 21 social-building opportunities, including bowling, sport and societies and a final night party. HE info sessions include those on student support, future plans, careers, myth busting, student finance, and a session for parents and carers. Pupils pre-select their six subject tasters from a selection of between two-three available simultaneously. Content is delivered by the relevant expert: academic lecturers, student support service staff, student ambassadors and outreach practitioners.

University B

These summer schools are for post-16 aged students and vary by specific subject as outlined below. Activities delivered across all summer schools include:

- Subject specific lectures and taster sessions;
- Interactive workshops/tutorials/demos delivered by academic staff and student ambassadors to expand subject knowledge;
- Talks to explain the application and admissions process;
- Careers talks and/or employability sessions to explain the benefits of choosing particular subjects at UG level;
- Activities to foster a sense of belonging with the university;
- Team building activities to encourage engagement with the event and each other;
- Practical activities to support application to HE such as personal statement workshop, how to choose a course/university, contextual admissions scheme and financial support information;
- Information sessions about campus and accommodation;
- Q&A with current undergraduates.

Biosciences

This summer school includes two days online, and one day on campus. The online sessions are made up of academic sessions, social time, workshops on careers and employability, and pre-recorded sessions available throughout such as a virtual campus tours and academic lectures. The on-campus activities include ice breakers, lab workshops, a campus tour and motivational speaker.

Dentistry

A two-day on campus summer school with one night staying in university accommodation. Sessions include welcome and icebreakers, first year taster lecture, campus tour, clinical skills activity, applying to dentistry – information session and Q and A with current medical students, communication skills and ethics in a dentistry setting, learning how to make judgements and decisions, admissions

test session, personal statement workshop and general Q and A with staff and students.

Medicine
A two-day on campus summer school with one night staying in university accommodation. Sessions include welcome and icebreakers, first year taster lecture, campus tour, clinical skills activity, tips and strategies for applying to medicine and Q and A with current medical students, communication skills and medical ethics, learning how to make judgements and decisions, admissions test session, personal statement workshop and general Q and A with staff and students.

Psychology
This summer school includes two days online, and one day on campus. The online sessions are made up of academic tasters, life as a Psychology student, social time, and workshops on careers and employability. The on-campus activities include academic lectures, lab workshops, a campus tour and Q and A with student ambassadors.

Social Sciences
A two-day on campus summer school with one night staying in university accommodation. Activities include a welcome and ice-breaker session, campus tour, 5 x 1 hour workshops on subjects and student life, presentation planning and delivery, reflection time and a social activity on campus.

University C
These summer schools are for post-16 aged students and vary by specific subject as outlined below. All take place on campus over four days, with three nights spent in university accommodation.

Arts
Students on the Arts stream have sessions including a campus tour, a welcome talk, a project overview, Adapting Shakespeare introductory talk, clips and discussion, and a film and how it works talk. Further workshops include A Cultural History of Romeo and Juliet in Cinema taster lecture, and From Pages to Screen group work and filming (for their project). On the final day there are three sessions; Viewing films and reflection, what does Shakespeare's work look like in foreign language film adaptations and a Q&A. A total of 12 subject specific sessions.

Business
Students on the Business stream have sessions including a campus tour, Innovation lecture, Innovation Group work, Business Ethics lecture and Business ethics groupwork. Further activities include a financial markets talk and a supply chain talk with additional sessions on sales and pricing, and a studying at the school of business Q&A. A total of 11 subject specific sessions.

Law
Students on the Law stream have sessions on a crime scene, why study law, a campus tour and on homicide and interviewing clients. Further workshops include interviewing and advising, plea in mitigation, presenting your plea in mitigation, impact of imprisonment and Q&A. A total of 9 subject specific sessions.

Medicine
Students on the Medicine stream have sessions including a campus tour, a working in the NHS talk, a taster lecture about strokes and a group activity on a patient journey regarding strokes. Further workshops include a multidisciplinary management of stroke lecture, group work analysing patient notes, and a University Clinical Aptitude Test (UCAT)/personal statement preparation session with an additional optional session on UCAT practice questions. Final day sessions were on multi-mini-interviews, and applying to medicine and healthcare courses. A total of 9 subject specific sessions.

STEM
Students on the STEM stream have sessions including a chemistry chlorophyll practical, a geology/geography planetary atmospheres and life lecture, a Life Science – DNA and Microbes practical, and a Natural Sciences – Astrobiology: the possibility of life beyond Earth lecture. Further workshops include a Life Sciences – checking plates for bacterial growth practical, a Life Sciences – Mutants under the microscope practical, a campus tour and a mentor Q&A. A total of 8 subject specific sessions.

University D pre-16 and post-16 summer schools
A four-day on-campus summer school with three nights spent in university accommodation. Separate summer schools are run for Year 10 and Year 12 students with both exploring the theme of ‘Breaking Barriers’ (though pitched at different levels), encouraging participants to join the university’s pledge to build a fairer world. Participants will have the chance to experience what it is like to be at university, experiencing different aspects of student life, from cooking to participating in sports and social activities and making new friends. Alongside this, participants will explore how learning happens at university and will build their own skills through the Breaking Barriers activities based around personal barriers, academic barriers and building a fairer community.

University E
This summer school is a two-day on campus summer school with one night staying in university accommodation. The summer school is for Year 9 students designed to give pupils an insight into what university life could be like. They will get to meet and work with pupils from other schools and experience a range of sessions, including those on university life, subject tasters, student finance, clubs and societies and a Q&A with student ambassadors.

Appendix II: Distribution of responses to outcome survey questions

	Treatment (N = 661)	Control (N = 1,026)
Likelihood of going to HE		
Extremely likely	224 (33.9%)	138 (13.5%)
Likely	66 (10.0%)	51 (5.0%)
Somewhat likely	27 (4.1%)	15 (1.5%)
Neutral	19 (2.9%)	4 (0.4%)
Somewhat unlikely	2 (0.3%)	0 (0%)
Unlikely	1 (0.2%)	3 (0.3%)
Extremely unlikely	0 (0.0%)	1 (0.1%)
Missing	322 (48.7%)	814 (79.3%)
Likelihood of progressing to academic study post-16		
Extremely likely	49 (7.4%)	38 (3.7%)
Likely	18 (2.7%)	8 (0.8%)
Neutral	10 (1.5%)	1 (0.1%)
Unlikely	1 (0.2%)	3 (0.3%)
Extremely unlikely	0 (0%)	0 (0%)
Missing	583 (88.2%)	976 (95.1%)
Self-efficacy relating to HE		
Extremely confident	78 (11.8%)	45 (4.4%)
Quite confident	191 (28.9%)	127 (12.4%)
Neutral	54 (8.2%)	29 (2.8%)
Not that confident	11 (1.7%)	6 (0.6%)
Not confident at all	3 (0.5%)	1 (0.1%)
Missing	324 (49.0%)	818 (79.7%)
Compatibility of HE with social identity		
Strongly agree	85 (12.9%)	48 (4.7%)
Agree	133 (20.1%)	75 (7.3%)
Neither agree nor disagree	85 (12.9%)	68 (6.6%)
Disagree	12 (1.8%)	16 (1.6%)

Strongly disagree	4 (0.6%)	2 (0.2%)
Missing	342 (51.7%)	817 (79.6%)
Perception of practical barriers to HE		
Extremely confident	36 (5.4%)	15 (1.5%)
Quite confident	83 (12.6%)	52 (5.1%)
Neutral	143 (21.6%)	78 (7.6%)
Not that confident	59 (8.9%)	53 (5.2%)
Not confident at all	14 (2.1%)	10 (1.0%)
Missing	326 (49.3%)	818 (79.7%)
Applied to HE		
Yes	87 (13.2%)	102 (9.9%)
No	9 (1.4%)	8 (0.8%)
Missing	565 (85.5%)	916 (89.3%)
Notes: Total n represents the randomised sample.		

Table 14: Distribution of responses to outcome survey questions

Appendix III: Eligibility criteria, guaranteed places and quotas specified by HEPs

HEP	Pre or post-16	Delivery mode	Subject specific?	Guaranteed places	Eligibility criteria	Quotas
University D pre-16	Pre-16	Residential	No	Care-experienced	Attend partner school/college, one WP criteria - first-gen, BAME, POLAR 4 Q1, IMD Q1, disability	No quota (but only space for 33 males in accommodation so upper bound for treatment group)
University E	Pre-16	Residential	No	No	One of the following - IMD Q1, POLAR4 Q1, FSM, disability, care-experienced, young carer	At least 5 females and 5 males from each of the 9 partner schools (where enough applicants)
University A	Pre-16	Residential	No	No	One of the following - first-gen, FSM, IMD Q1, POLAR4 Q1, disability or BAME	At least 1 student from each of the 12 partner schools, and at least 25 males to be allocated a place
University B (Dentistry)	Post-16	Residential	NA	Males if only a small number apply	One of the following - low participation neighbourhood, FSM, low income (£25,000 per annum or less), care-experienced or studies disrupted (studies disrupted by circumstances in their personal, social or domestic lives).	12 males applied and all must be allocated a place. 51 females applied, only had room for 16 to get a place
University B (Medicine)	Post-16	Residential	NA	No		40 males applied, only had room for 28 males to get a place. 120 females applied, only had room for 16 to get a place
University B (Psychology)	Post-16	Offering online and face-to-face	NA	No		Only 5 males applied, and wanted 3 to be guaranteed a place
University B (Social Sciences)	Post-16	Residential	NA	No		NA
University B (Biosciences)	Post-16	Offering online and face-to-face	NA	Identify as Gypsy, Roma, or Traveller		NA
University C (Medicine)	Post-16	Residential	NA	Care-experienced and low participation areas - approx. 10	One of the following - care-experienced, young carer, disability, estranged, FSM, GRT, refugee or asylum seeker, first-gen, POLAR4 Q1	NA

University C (Law)	Post-16	Residential	NA	Care-experienced		NA
University C (Arts)	Post-16	Residential	NA	Care-experienced		NA
University C (STEM)	Post-16	Residential	NA	Care-experienced		NA
University C (Business)	Post-16	Residential	NA	Care-experienced		NA
University D Post-16	Post-16	Residential	NA	Care-experienced	Attend partner school/college, one WP criteria - first-gen, BAME, POLAR 4 Q1, IMD Q1, disability	NA

Table 15: Eligibility criteria, guaranteed places and quotas specified by HEPs

Appendix IV: Full regression tables for the main pre-specified model (Model 1)

	Outcome variable:		
	Likelihood of going to HE	Likelihood of progressing to study post-16	Self-efficacy relating to HE
	(1)	(2)	(3)
Treatment	-0.05 (0.17)	-0.25 (0.26)	-0.10 (0.14)
Sex: Male	-0.05 (0.22)	-0.23 (0.37)	0.05 (0.20)
Sex: Other/Unknown	-0.22 (0.67)	-0.53 (0.61)	-0.40 (0.48)
Ethnicity: Asian	0.07 (0.25)	-0.33 (0.45)	0.04 (0.22)
Ethnicity: Missing	-0.46 (0.41)	-1.90*** (0.25)	-0.002 (0.56)
Ethnicity: Other	0.13 (0.33)	0.42 (0.37)	0.49** (0.20)
Ethnicity: Black	-0.21 (0.28)	0.35 (0.34)	0.28 (0.22)
IMD: Quintile 2	-0.12 (0.25)	0.21 (0.33)	0.06 (0.18)
IMD: Quintile 3	-0.004 (0.23)	0.27 (0.39)	-0.06 (0.17)
IMD: Quintile 4	0.03 (0.40)	0.64* (0.38)	-0.27 (0.26)
IMD: Quintile 5	0.13 (0.35)	0.29 (0.41)	-0.26 (0.30)
POLAR4: Quintile 2	-0.04 (0.23)	-0.60 (0.40)	-0.18 (0.18)
POLAR 4: Quintile 3	-0.22 (0.24)	-0.72** (0.35)	0.14 (0.19)
POLAR4: Quintile 4	-0.18 (0.32)	-0.38 (0.57)	0.24 (0.21)
POLAR4: Quintile 5	0.05 (0.31)	-0.12 (0.45)	0.09 (0.22)
FSM: Yes	-0.02 (0.20)	-0.18 (0.25)	-0.22 (0.16)
FSM: Unknown	-0.09 (0.37)		0.37 (0.36)
Parent attended HE: Yes	-0.13 (0.18)	-0.10 (0.25)	-0.13 (0.14)

Parent attended HE: Don't know	-0.09 (0.81)	-0.18 (0.71)	-0.62 (0.61)
Intercept	6.62*** (0.60)	5.85*** (0.74)	4.94*** (0.57)
Summer school ID	Yes	Yes	Yes
Randomisation block ID	Yes	Yes	Yes
School ID	Yes	Yes	Yes
N. obs	541	119	536

Note: $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$
For categorical variables with over 10 values (summer school, randomisation block, school), we don't report the full results so the table remains legible. "Yes" means the variable is included in the regression.

	Outcome variable:			
	Compatibility of HE with social identity	Perception of practical barriers to HE	Perception of financial barriers to HE	Perception of knowledge barriers to HE
	(1)	(2)	(3)	(4)
Treatment	0.31^{+} (0.16)	0.01 (0.18)	0.03 (0.18)	0.02 (0.17)
Sex: Male	0.19 (0.24)	0.27 (0.26)	0.27 (0.26)	-0.01 (0.25)
Sex: Other/Unknown	-0.25 (0.41)	-0.33 (0.54)	-0.33 (0.54)	0.29 (0.57)
Ethnicity: Asian	-0.32 (0.26)	0.36 (0.26)	0.36 (0.26)	0.31 (0.26)
Ethnicity: Missing	-0.21 (0.46)	0.77 (0.86)	0.77 (0.86)	-0.34 (0.79)
Ethnicity: Other	0.02 (0.25)	0.59* (0.30)	0.59* (0.30)	0.20 (0.28)
Ethnicity: Black	-0.08 (0.26)	0.48^{+} (0.26)	0.47^{+} (0.25)	0.16 (0.27)
IMD: Quintile 2	0.14 (0.21)	-0.02 (0.21)	-0.02 (0.21)	0.17 (0.22)
IMD: Quintile 3	0.12 (0.22)	0.01 (0.26)	0.01 (0.26)	0.20 (0.24)
IMD: Quintile 4	-0.50 (0.32)	0.03 (0.34)	0.04 (0.34)	-0.13 (0.35)
IMD: Quintile 5	0.16 (0.37)	-0.005 (0.38)	-0.01 (0.38)	-0.25 (0.38)
POLAR4: Quintile 2	0.10 (0.22)	-0.15 (0.25)	-0.14 (0.25)	0.05 (0.24)
POLAR 4: Quintile 3	0.35 (0.26)	-0.33 (0.25)	-0.32 (0.25)	-0.12 (0.26)
POLAR4: Quintile 4	0.60* (0.31)	0.57^{+} (0.34)	0.58^{+} (0.34)	0.13 (0.30)
POLAR4: Quintile 5	0.28 (0.33)	0.16 (0.40)	0.17 (0.40)	0.16 (0.35)
FSM: Yes	-0.02 (0.19)	-0.06 (0.20)	-0.07 (0.20)	-0.08 (0.20)
FSM: Unknown	0.20 (0.41)	0.21 (0.35)	0.20 (0.35)	0.09 (0.38)

Parent attended HE: No	-0.18 (0.16)	-0.29^{+} (0.17)	-0.30^{+} (0.17)	-0.19 (0.16)
Parent attended HE: Don't know	-0.96 (0.63)	-0.59 (0.65)	-0.60 (0.65)	0.44 (0.61)
Intercept	4.72** (0.46)	3.05** (0.93)	3.03** (0.92)	3.30** (0.95)
Summer school ID	Yes	Yes	Yes	Yes
Randomisation block ID	Yes	Yes	Yes	Yes
School ID	Yes	Yes	Yes	Yes
N. obs	519	534	537	539

Note: $^{+}$p<0.1; *p<0.05; **p<0.01
For categorical variables with over 10 values (summer school, randomisation block, school), we don't report the full results so the table remains legible. "Yes" means the variable is included in the regression.

Appendix V: Full regression table for the linear estimation model for the outcome 'Applied to HE'

	Outcome variable: Applied to HE
Treatment	-0.14 (0.23)
Sex: Male	-0.10 (0.23)
Sex: Other/Unknown	0.19 (0.73)
Ethnicity: Asian	0.12 (0.25)
Ethnicity: Other	0.18 (0.28)
Ethnicity: Black	0.24 (0.26)
IMD: Quintile 2	0.13 (0.25)
IMD: Quintile 3	0.20 (0.41)
IMD: Quintile 4	0.17 (0.34)
IMD: Quintile 5	0.11 (0.25)
POLAR4: Quintile 2	-0.06 (0.33)
POLAR 4: Quintile 3	-0.12 (0.33)
POLAR4: Quintile 4	-0.17 (0.33)
POLAR4: Quintile 5	-0.27 (0.37)
FSM: Yes	0.04 (0.20)
FSM: Unknown	0.14 (0.31)
Parent attended HE: No	0.03 (0.18)
Intercept	0.95* (0.40)
Summer school ID	Yes
Randomisation block ID	Yes
School ID	Yes
N. obs	203

Note: +p<0.1; *p<0.05; **p<0.01

For categorical variables with over 10 values (summer school, randomisation block, school), we don't report the full results so the table remains legible. "Yes" means the variable is included in the regression.